ELEMENTARY

Differential Equations

ELEMENTARY
Differential Equations

By

THORNTON C. FRY, Ph.D.
Member of the Technical Staff
Bell Telephone Laboratories, Inc.

THIRD PRINTING

NEW YORK
D. VAN NOSTRAND COMPANY, Inc.
250 FOURTH AVENUE

First Published, February, 1929
Second Printing, December, 1929
Third Printing, October, 1930

Printed in U. S. A.

PRESS OF
BRAUNWORTH & CO., INC.
BOOK MANUFACTURERS
BROOKLYN, N. Y.

PREFACE

This text, which has evolved through several mimeographed editions in the Out-of-Hour Courses of Bell Telephone Laboratories, is intended primarily for students of Engineering and its allied sciences.

To such an audience technical applications and illustrations are of the highest value, for they stimulate a lively interest in, and add depth of meaning to, abstract mathematical ideas which otherwise might remain rather vague. They are not without their danger, however, if they are used unwisely, for should the student come to rely upon his physical intuition as a substitute for abstract logic, rather than as an aid to it, he would have missed half the benefit of his study. He might indeed acquire a considerable degree of proficiency in operating the machinery of mathematics, but he would lack that deeper understanding of principles which forms the dividing line between mechanic and engineer.

In order to avoid this potential danger, most of the illustrative material in this text has been collected into separate chapters. Not only does this arrangement allow an uninterrupted development of mathematical ideas, but it also permits the inclusion of many more examples than would otherwise be feasible. The freedom of choice provided by this excess of material may perhaps be of little value to the instructor, who will probably draw his illustrations largely from the subject in which he is at the moment most interested, whether it be in the text or not; but the Out-of-Hour courses have shown it to have another sort of merit, in that the better students develop a spontaneous interest in these illustrations, even when they are not assigned, and follow them up on their own initiative. The educational value of such voluntary effort needs no special emphasis.

In the final revision of the text I have been much benefited by the advice of Mr. L. A. MacColl and Mr. G. G. Muller, of Bell Telephone Laboratories, both of whom have conducted Out-of-Hour Courses on the subject, and of Professor J. H. M. Wedderburn, of Princeton University, who read the manuscript and made a number of valuable suggestions. I have also been aided in many ways by Miss Clara L. Froelich, of Bell Telephone Laboratories, who, among other things, prepared the figures with which the book is illustrated, furnished answers to the problems and read the proof sheets.

Thornton C. Fry.

New York,
February 1, 1929.

CONTENTS

CHAPTER I

Introduction

CHAPTER II

The Significance of Differential Equations and Their Solutions

CHAPTER III

The Origin of Differential Equations

CHAPTER IV

Methods of Solution; First Order Equations

CHAPTER V

Singular Solutions

CHAPTER VI

Practical Applications of Differential Equations

CHAPTER VII

Linear Equations of Order Higher than the First

CHAPTER VIII

Systems of Linear Equations

CHAPTER IX

Other Equations of Order Higher than the First

CHAPTER I

Introduction

§ 1. *Preliminary Definitions*

Any function of a set of variables and their derivatives is a *differential expression*, and any equation containing a differential expression is a *differential equation*. If an equation has derivatives with respect to one variable only it is called an *ordinary differential equation;* otherwise it is a *partial differential equation*. In either case the variables with respect to which differentiation is performed are *independent variables*, and those of which the derivatives are taken, *dependent variables*. The last two terms, however, although they are indispensable in speaking and writing about differential equations, represent a somewhat artificial distinction; for it is always possible to so transform a differential expression as to cause some of the variables which originally were independent to appear as dependent variables, and *vice versa*.

The *order* of a differential equation is the order of the highest derivative involved. Thus the order of the equation

$$\frac{\partial^3 \phi}{\partial x^3} + 2\frac{\partial \phi}{\partial x} + \frac{\partial^2 \phi}{\partial y^2} + \left(\frac{\partial \phi}{\partial y}\right)^4 = 0$$

is 3.

The *degree* of a differential equation is the exponent of the derivative of highest order appearing in the equation after it is completely rationalized and cleared of fractions. That is, the degree of the equation above is 1, because the exponent of $\dfrac{\partial^3 \phi}{\partial x^3}$ is 1. On the other hand, the degree of

$$\frac{d^2 y}{dx^2} = \sqrt{1 + y^2 + \frac{dy}{dx}}$$

ing, it would lead to the conclusion that (4) has no solution. This conclusion would result, not from the fact that (5) does not adequately define y as a function of x, but from the fact that there are integrals which cannot be expressed in the form demanded by algebra; and this in turn is due to the more or less arbitrary choice of those functions which are termed "elementary." On the whole, it appears wiser to extend the term "solution" to cover any functional relation between the dependent and independent variables (but not containing their derivatives) which causes the differential equation to be identically satisfied. This is the content of the definition given above.

§ 2. *Change of Variable in Differential Expressions*

Many differential expressions may be greatly simplified by replacing one of the variables by some function of one or more of them. It not infrequently occurs that when an expression has been so transformed it can be integrated immediately by inspection or by some method of integration previously developed. In case the differential expression occurs in a differential equation this process may lead to a solution of the equation: as a matter of fact, many important "standard methods of solution" can be so interpreted. Hence the study of differential equations may properly be begun by reviewing briefly the technique of change of variable.

Consider an arbitrary function of two variables x and y and the successive x-derivatives of y. We may denote it by

$$F(x, y, y', y'', \cdots),$$

$y', y'', \cdots$, being written briefly for $\frac{dy}{dx}, \frac{d^2y}{dx^2}, \cdots$. No matter how complicated this expression may be, it can be transformed by introducing a new variable in place of one of the old ones, provided each of the arguments $x, y, y', y'', \cdots$, can itself be so transformed.

Suppose it is the dependent variable y which is to be

replaced, and that the new variable w which is to supplant it is related to x and y by a known equation

$$f(x, y, w) = 0. \tag{6}$$

The process in this case is quite simple. It is only necessary to solve (6) for y, so as to obtain a relation of the form

$$y = \phi(x, w),$$

and then to replace $y, y', y'', \cdots$, by ϕ and the expressions that are obtained from ϕ by repeated differentiation with respect to x. As ϕ does not contain y none of its x-derivatives will contain y, and therefore no y's will remain in the differential expression after the substitutions have been made.

As an example, consider the differential expression

$$x^2 \frac{d^2y}{dx^2} + 4x \frac{dy}{dx} + 2y. \tag{7}$$

To replace y by a new variable w, related to x and y by the equation

$$x^2y - w = 0,$$

it is only necessary to find y and its first two derivatives in terms of x and w. But

$$y = \frac{w}{x^2}, \tag{8}$$

and hence by actual differentiation

$$\frac{dy}{dx} = \frac{1}{x^2}\frac{dw}{dx} - \frac{2w}{x^3},$$

$$\frac{d^2y}{dx^2} = \frac{1}{x^2}\frac{d^2w}{dx^2} - \frac{4}{x^3}\frac{dw}{dx} + \frac{6}{x^4}w.$$

Substituting these values in (7) the entire expression reduces to $\dfrac{d^2w}{dx^2}$.

In case the *expression* (7) is equated to zero there results the *equation*

$$x^2 \frac{d^2y}{dx^2} + 4x \frac{dy}{dx} + 2y = 0. \tag{9}$$

If the student were confronted with the necessity of solving this equation and possessed only the tools furnished by the Integral Calculus he might be at a loss how to proceed; but once the substitution (8) has been made the solution becomes evident at once. In fact, in terms of w (9) takes the form

$$\frac{d^2w}{dx^2} = 0,$$

the solution of which is known to be

$$w = \alpha x + \beta, \tag{10}$$

α and β being any constants whatsoever. By replacing w in (10) by its value x^2y, the solution of (9) is obtained at once in the form

$$y = \frac{\alpha}{x} + \frac{\beta}{x^2}.$$

If it is desired to introduce the new variable w in place of the *independent* variable x, the procedure is somewhat more complicated, although it is still perfectly straightforward. Let (6) be solved for x, the solution being

$$x = \psi(y, w). \tag{11}$$

Differentiation of this expression with respect to w gives

$$\frac{dx}{dw} = \frac{\partial\psi}{\partial y}\frac{dy}{dw} + \frac{\partial\psi}{\partial w}.$$

This entire quantity is a function of the three arguments y, w and $\frac{dy}{dw}$ only, for $\frac{\partial\psi}{\partial y}$ and $\frac{\partial\psi}{\partial w}$ do not contain x. Let it be denoted by f for simplicity in writing. Then $dx = f\,dw$, and therefore

$$y' = \frac{1}{f}\frac{dy}{dw}. \tag{12}$$

The right-hand member of (12) does not explicitly contain x, hence (11) and (12) together are sufficient to eliminate x from any differential expression which does not contain

derivatives of y higher than the first. If higher derivatives do occur it is only necessary to continue the same processes. Thus the quantity which must replace y'' may be found by noting that

$$y'' = \frac{d}{dx}y' = \frac{1}{f}\frac{d}{dw}y' = \frac{1}{f}\frac{d}{dw}\left(\frac{1}{f}\frac{dy}{dw}\right).$$

In general

$$y^{(n)} = \frac{1}{f}\frac{d}{dw}\frac{1}{f}\frac{d}{dw}\cdots\frac{1}{f}\frac{dy}{dw}, \tag{13}$$

the symbol $\frac{1}{f}\frac{d}{dw}$ being repeated n times in succession. This expression is often written in the shorthand form

$$y^{(n)} = \left(\frac{1}{f}\frac{d}{dw}\right)^n y.$$

Though (12) and (13) may appear formidable, the processes which they represent are often extremely simple. As formulæ they are of no importance, for once the method is properly understood it is generally simpler and quicker to carry out the required operations than to apply the formulæ themselves.

For instance, consider again the expression (7), and let

$$x = e^w \tag{11'}$$

be the relation between x, y and w which corresponds to equation (6) of the general theory; just as $x^2y - w = 0$ did in the preceding example. As it is already solved for x, it also corresponds to (11). Its w-derivative is

$$\frac{dx}{dw} = e^w,$$

or, in the terms of the general discussion, $dx = f\,dw$, the f in this case meaning e^w. Hence

$$y' = e^{-w}\frac{dy}{dw}, \tag{12'}$$

and

$$y'' = e^{-w}\frac{d}{dw}e^{-w}\frac{dy}{dw} = e^{-2w}\left(\frac{d^2y}{dw^2} - \frac{dy}{dw}\right). \tag{13'}$$

With the aid of these, (7) becomes

$$\frac{d^2y}{dw^2} + 3\frac{dy}{dw} + 2y, \tag{7'}$$

the form of which differs from (7) in that the coefficients of the derivatives are all constant.

A series of problems assigned in this and later chapters deals with the type of expression to which (7) belongs; that is, with Cauchy linear equations. Collectively they show that such equations can always be reduced to the form (7′), in which all the coefficients are constant, and that whenever the latter type can be solved, the former can also.

§ 3. *Continuity* [1]

The curve of Fig. 1 is said to be discontinuous at A, but continuous at B. That these terms are appropriate is obvious

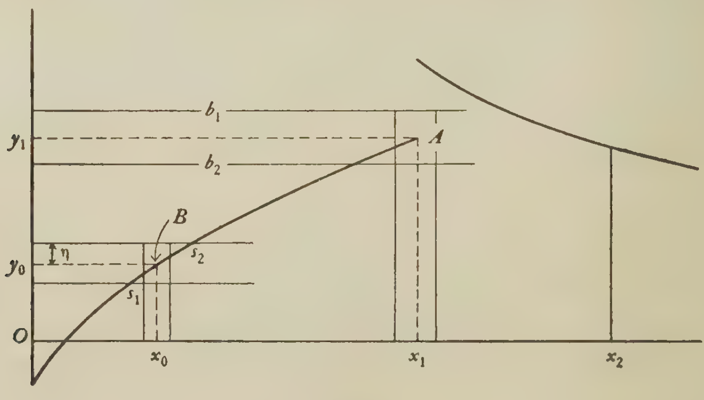

FIG. 1.

from a common-sense point of view, but a more exact formulation of their meanings is necessary before they can be used for mathematical purposes.

Consider first the point B, the coordinates of which are

[1] The subjects discussed in the remaining sections of this chapter are only indirectly related to differential equations. They are collected together in this place so that it may not be necessary to interrupt the main argument later on, when we find it necessary to refer to them.

x_0 and y_0, and draw a pair of horizontal lines at the heights $y_0 + \eta$ and $y_0 - \eta$. These cut off a short segment s_1s_2 of the curve. It is obvious that a pair of vertical lines can be drawn, one between x_0 and s_2, the other between s_1 and x_0, in such a way that in the interval between them the curve lies constantly between the horizontals $y_0 + \eta$ and $y_0 - \eta$. If a new value of η is chosen, smaller than the old one, s_1 and s_2 will move nearer together, and it may be necessary to move the vertical lines closer together also; but the statement will still remain true, that there is some finite interval about x_0 within which the curve nowhere passes outside the pair of horizontal lines.

This statement is not true in the case of the point A, however, if the horizontals are chosen as b_1 and b_2; for no matter how small the interval about x_1 may be made, the portion of the curve to the right of A lies outside the pair of horizontals.

Now, obviously the reason the statement is not true of A is, that the curve is not continuous at A; and this leads us to adopt the following definition of a continuous curve:

A curve is said to be continuous at x_0 *if it is possible, upon assigning a value* η, *to find an interval about* x_0 *within which the ordinate nowhere deviates from its height at* x_0 *by a greater amount than* η, *no matter how small* η *may be.*

The same idea can be phrased in terms of the function which the curve represents. It then takes the form: *y is a continuous function of x at $x = x_0$ if, when a number η is assigned, a number ϵ can be found such that in the interval between $x_0 - \epsilon$ and $x_0 + \epsilon$ the value of y remains constantly between the limits $y_0 - \eta$ and $y_0 + \eta$, no matter how small η may be.* If this statement is violated the function is discontinuous.

It will be noticed that continuity is defined as a property, not of the curve as a whole, but of individual points on it. It is a simple matter, however, to extend the idea somewhat and say that *if a curve is continuous at every point of the segment lying between the two abscissæ x' and x'', it is continuous between these limits.* Similarly, the function $f(x)$ represented by it is

continuous in the interval $x'x''$. Thus the curve of Fig. 1 is continuous between 0 and x_1 and between x_1 and x_2. But it is not continuous between 0 and x_2, for it is discontinuous at the point x_1 which lies in this interval.

§ 4. *Differentiability*

In his study of calculus the student has been concerned largely with the formal processes by means of which functions are differentiated, and has generally proceeded upon the assumption that a derivative exists and may be found by proper manipulation. The assumption that a derivative exists, however, is not always justified, as may be shown by citing an example in which there is none.

The function $y = x \sin \frac{1}{x}$ is represented by the curve of Fig. 2. Its value at $x = 0$ is somewhat uncertain, but may be arbitrarily defined to be zero, which makes the function continuous. To find the derivative at this point, x must be caused to take an increment Δx, and the increment Δy which the function takes on in consequence must be noted. These values, as shown in the figure, locate a point A. As Δx becomes smaller and smaller A moves inward toward the origin, following the course of the curve. By definition the desired derivative is the limit of $\frac{\Delta y}{\Delta x}$ as Δx is made to approach 0. But $\frac{\Delta y}{\Delta x}$ is the slope of the line OA, which oscillates back and forth between the lines OT_1 and OT_2 as Δx decreases. No matter

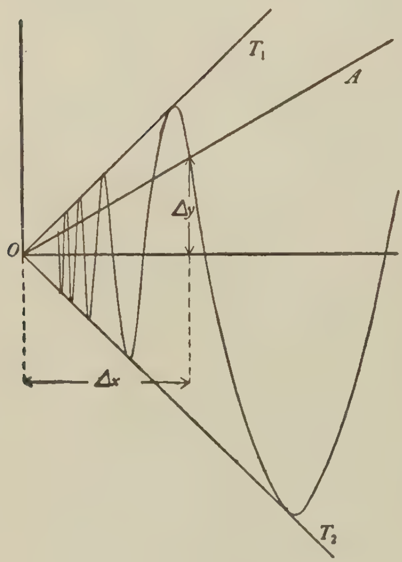

Fig. 2.

how small Δx becomes this oscillation continues; for no matter how small the distance to the origin may be there are points included in it at which the curve is tangent to OT_1, and others at which it is tangent to OT_2. Therefore the slope of OA passes from $+1$ to -1 and back again (these are the slopes of OT_1 and OT_2), repeating the process time after time and never approaching any limit at all. Hence there is no limit to $\frac{\Delta y}{\Delta x}$ as Δx approaches 0: that is, no value can be found for $\frac{dy}{dx}$ at the origin.

If an attempt is made to find the derivative at any other point of the curve, the secant line ultimately stops oscillating as Δx decreases and in the end approaches as a limit the tangent line to the curve. Hence the function has a derivative for every value of x except $x = 0$; but at this latter value there is no derivative.

It is therefore evident that there are functions which are incapable of differentiation. In § 11 it will be found that because of this certain problems in differential equations cannot be solved.

§ 5. *Envelopes*

Consider an equation of the form $f(x, y, c) = 0$, in which c is a constant the value of which can be assigned at pleasure. For each value of this constant the function defines a curve. The equation therefore represents a family of curves.

There are certain properties of this family of curves which may be stated at once. For example, the number of curves passing through any point (x_1, y_1) is equal to the number of values of c which satisfy the equation $f(x_1, y_1, c) = 0$. If this equation is of the nth degree in c it will have n roots. For simplicity, we shall call these "the c's belonging to (x_1, y_1)." To each of these c's corresponds a curve passing through the point (x_1, y_1).

It will generally be true that all these roots of the equation

are distinct. If so the n curves are all different members of the family. In exceptional cases, however, the equation may have equal roots, and then the same curve must be counted twice. Naturally, these exceptional cases are due to something unusual about the family of curves. Our purpose in the present section is to study these exceptional cases, some of which are of importance in the study of differential equations. Before we can proceed, however, we must have some method of sorting out the particular points at which equal c's occur.

To derive such a method, we shall suppose that any pair of values (x_1, y_1) are chosen, without any reference to whether they give equal c's or not. We can then plot the function $f(x_1, y_1, c)$ as a function of c, obtaining in this fashion some such curve as that shown in Fig. 3.

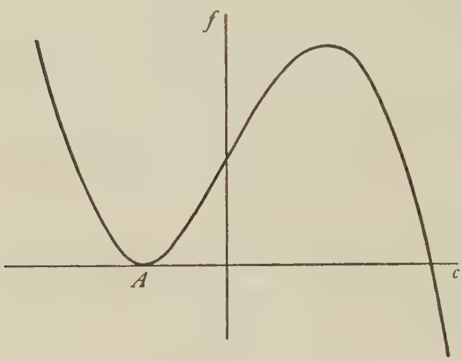

Fig. 3.

Wherever this curve crosses the axis the relation $f(x_1, y_1, c) = 0$ is satisfied: that is, the curve of Fig. 3 crosses the axis at just those c's which belong to (x_1, y_1). These c's are all distinct unless, as at the point A of Fig. 3, the curve happens to be *tangent* to the axis; then multiple roots appear. But if the curve is tangent to the axis it is true, not only that

$$f(x_1, y_1, c) = 0,$$

but also that

$$\frac{\partial}{\partial c} f(x_1, y_1, c) = 0$$

as well.

By eliminating c from these two equations we may get a relation between x_1 and y_1 which must be satisfied wherever equal c's appear. Let us call it $F(x_1, y_1) = 0$. Then we can say at once that equal c's cannot occur at the point (x_1, y_1)

unless it lies upon the curve $F(x, y) = 0$, the equation of which is found by eliminating c between the equation $f(x, y, c) = 0$ which defines our family of curves, and the equation $\frac{\partial}{\partial c}f(x, y, c) = 0$ which results when $f = 0$ is differentiated with respect to c.

For example, consider the family of curves defined by

$$(y - c)^2 - (x + c)^3 = 0. \tag{14}$$

As this is of the third degree in c, there are in general *three* different curves of the family passing through any point (x_1, y_1). To locate the exceptional points at which equal c's occur, we differentiate (14) with respect to c, getting

$$-2(y - c) - 3(x + c)^2 = 0. \tag{15}$$

From this it is found that

$$(y - c)^2 = \tfrac{9}{4}(x + c)^4; \tag{16}$$

and by combining this with (14) the new equation

$$[\tfrac{9}{4}(x + c) - 1](x + c)^3 = 0$$

is obtained. The solutions of this are $c = -x$ and $c = \frac{4}{9} - x$; and when substituted in (15) they give[1]

$$y + x = 0$$

and

$$y + x = \tfrac{4}{27},$$

respectively.[2]

It follows, then, that equal c's can only be found upon the pair of lines which these equations define.

[1] If the values of c had been substituted in (14) an additional equation $y + x = \frac{20}{27}$ would have been obtained. Though this is a part of the solution of (14) and (16), it is not a solution of (14) and (15) and therefore is of no consequence in our problem. It was brought into the result when (15) was squared to give (16).

[2] These *two* equations can be combined into *one*, which is then strictly analogous to the $F(x, y) = 0$ used above. For whenever either $y + x$ or $y + x - \frac{4}{27}$ is zero, their product is also; and conversely. Hence we can write the two together in the form

$$F(x,y) = (y + x)(y + x - \tfrac{4}{27}) = 0.$$

Having now found a method of isolating the points to which equal c's belong, we are prepared to return to our original question of determining what peculiarities of our family of curves may give rise to this unusual situation.

One is fairly obvious. If a curve crosses itself, as do those in the upper part of Fig. 4, it must be regarded as passing twice through the point of intersection. Equal c's must therefore occur at all such double-points.[1] Obviously, the truth of this statement is not affected by the size of the loops.

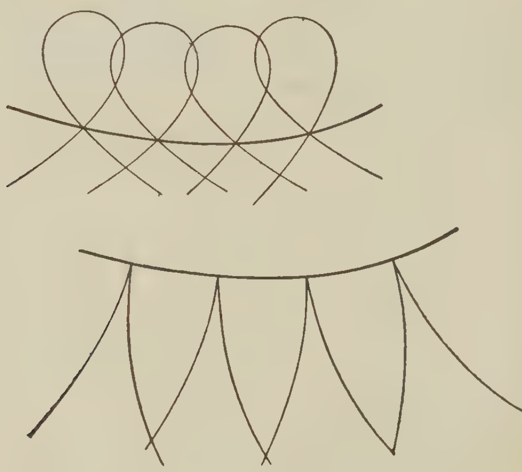

Fig. 4.—Illustrating the Nodal and Cuspidal Loci.

If, then, we allow them to shrink to nothing, thus producing the cusps shown in the lower part of the figure, we shall be led to expect equal c's at all such cusps.

These conclusions are immediately obvious. There is a third situation giving rise to equal c's, however, which is not so obvious, and as it is by far the most important from the standpoint of differential equations, we shall have to consider it in some detail. This situation arises when the curves of the family are so situated, as is the case with the family $a, d, c, b, \cdots$, of Fig. 5, that it is possible to draw in a curve A at every point of which one of the family is tangent. Such a curve is called an "envelope" of the family.

---

[1] A point at which a curve crosses itself is called a *double-point* or *node*.

For simplicity, we shall suppose the family to be defined by an equation $f(x, y, c) = 0$ of the second degree in c, so that there are only two curves through each point. We shall also center our attention upon a particular curve a, and shall note the point α at which it is tangent to the envelope A. Then let a set of points $\beta, \gamma, \delta, \cdots$, be chosen on a, each nearer to α than the preceding one. To each of these points, as to all other points of the plane, correspond two values of c: these determine the two curves which pass through the point.

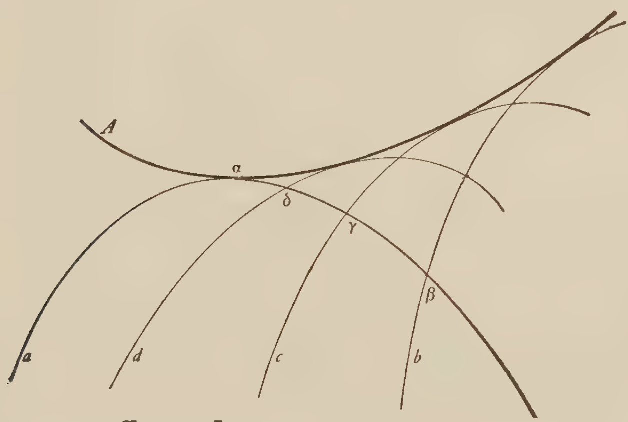

Fig. 5.—Illustrating the Envelope.

Whatever point is selected, one of these curves is obviously a itself. The other curves are different for different points. By drawing them in, a set of curves $b, c, d, \cdots$, is built up, each of which passes through one of the points $\beta, \gamma, \delta, \cdots$.

It is obvious that each of these curves is a nearer neighbor of a than the preceding one, and the c corresponding to it is likewise more nearly equal to the c of the curve a. In fact, by moving the point of intersection near enough to α, the neighboring curve can be brought as close as desired to a. In the limit, therefore, as the point of intersection becomes coincident with α, the curve selected by this process becomes identical with a: the two curves through α are therefore the same, and the two values of c are equal. Hence, to every point of the

envelope A correspond two equal values of c, which belong to that curve which is tangent to A at that point.

In résumé, therefore, it may be said that $f(x, y, c) = 0$ represents a family of curves, one for each value of c; and that upon eliminating c between this equation and the equation $\frac{\partial f}{\partial c} = 0$, we obtain the locus of all points to which belong equal values of c. This locus may include three types of curves:

1. Curves passing through all the double-points of the set $f(x, y, c) = 0$; these are called *nodal loci*;

2. Curves passing through all the cusps of the set $f(x, y, c) = 0$; these are called *cuspidal loci*;

3. Curves tangent at every point to some member of the set $f(x, y, c) = 0$; these are called *envelopes*.

One further point should be made: as these curves are loci of equal c's, they can only occur when at least two curves of the set pass through each point; and as the number of curves through a point is equal to the degree in c of the equation which defines the family of curves, it follows that a family cannot have an envelope unless its equation is of at least the second degree in c.

As a first example, consider the family of curves represented by the equation (14). They are obviously semi-cubical parabolas with their cusps located at $(-c, +c)$. As these points all lie on the line $x + y = 0$, this line is a cuspidal locus and should be obtainable by the processes just explained. We have already seen, however, that $x + y = 0$ is indeed one of the lines along which equal c's are to be expected. The other line of equal c's, $x + y = \frac{4}{27}$, is the envelope of the lower branches of the curves, as is obvious from Fig. 6.

As a second example, consider the equation

$$y^2 = (x - c)(x - 2c)^2. \tag{17}$$

Differentiating with respect to c and solving the resulting equation it is found that c must be $\frac{1}{2}x$ or $\frac{5}{6}x$.

The results of substituting these values in equation (17) are $y^2 = 0$ and $27y^2 - 2x^3 = 0$.

The locus of the first of these equations is the x-axis, which is a nodal locus. The other equation defines a semi-cubical

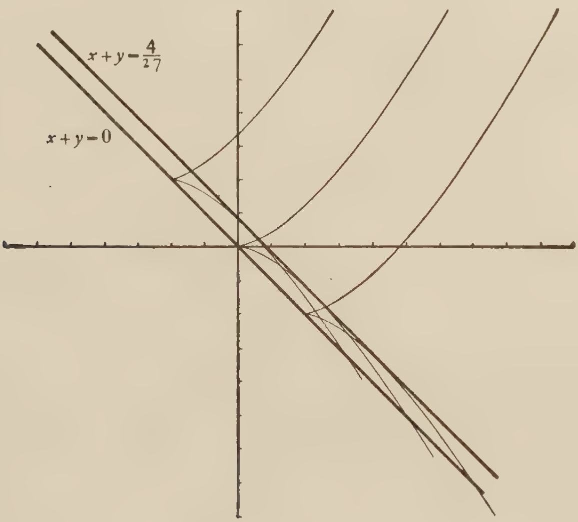

FIG. 6.—A FAMILY OF CURVES HAVING A CUSPIDAL LOCUS AND AN ENVELOPE.

parabola which is the envelope of the family of curves. The family is shown[1] in Fig. 7.

Finally, consider the family of circles

$$(x - c)^2 + y^2 - r^2 = 0, \tag{18}$$

all of which have the same radius r, but which have their centers distributed along the x-axis as shown in Fig. 8. It is obvious that the lines $y = r$ and $y = -r$ constitute the envelope and

[1] There is also another curve $16y^2 - x^3 = 0$, to which reference will be made later on.

that there are no cusps or double-points. Hence the solution of the equation (18) together with its c-derivative, should give us these two lines and nothing else. Differentiation with

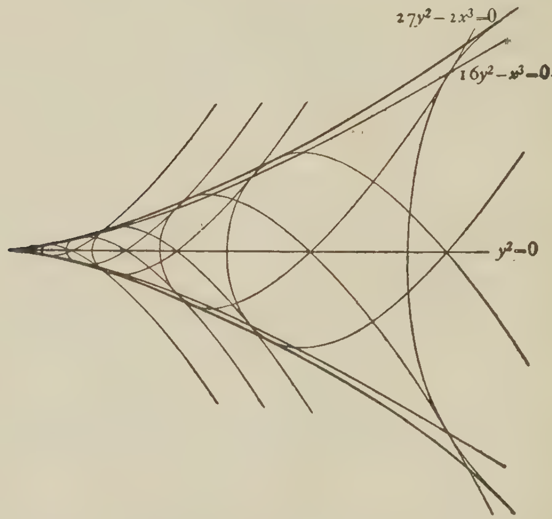

Fig. 7.—A Family of Curves Having a Tac Locus, a Nodal Locus and an Envelope.

respect to c yields $x - c = 0$, and substitution of this value in (18) gives $y^2 - r^2 = 0$. This is in fact the anticipated relation.

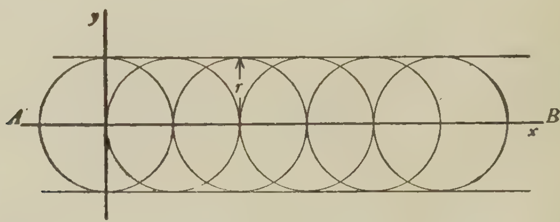

Fig. 8.—A Family of Curves Having a Tac Locus and an Envelope.

One final remark is necessary. We have spoken so far only of cases in which the number of distinct c's was finite. The main results of the argument, however, are equally valid

when the number of c's is infinite. Consider, for example, the family of sine curves

$$y - \sin cx = 0.$$

Through any point (x, y) there pass infinitely many of these curves. Yet the process of differentiating with respect to c and then eliminating c leads to the true envelope $y = \pm 1$.

PROBLEMS

1. Replace y in the expression

$$\frac{d^2y}{dx^2} + \frac{2}{x}\frac{dy}{dx} + y$$

by $w = xy$.

2. Replace y in the equation

$$x^2\frac{d^2y}{dx^2} - \frac{x^2}{2y}\left(\frac{dy}{dx}\right)^2 + 4x\frac{dy}{dx} + 4y = 0$$

by $w = \sqrt{y}$. Can you obtain the solution of this equation by comparing it with the first example in the text?

3. Replace y in the equation

$$\frac{dy}{dx} + cy = a$$

by $w = e^{cx}y$. What is the solution of this equation?

4. Replace r in the equation

$$\frac{d^2\theta}{dr^2} + \frac{1}{r}\frac{d\theta}{dr} + k^2\theta = 0 \qquad \text{[Bessel's Equation]}$$

by e^t.

5. Replace θ in the equation

$$\frac{d}{d\theta}\left(\sin\theta\frac{dP}{d\theta}\right) + mP\sin\theta = 0 \qquad \text{[Legendre's Equation]}$$

by $x = \cos\theta$.

6. Replace y in the equation

$$\frac{dy}{dx} = \frac{1}{x}\sqrt{1 - y^2}\sin^{-1} y$$

by $w = \sin^{-1} y$.

7. Differentiate (3) and solve for $\frac{dy}{dx}$. Show by substituting this value in (1) that (3) is really a solution of (1).

8. If $x = e^w$ and $z = e^{-nw}\frac{d^m y}{dw^m}$, find $\frac{dz}{dx}$ in terms of w and y.

9. Prove by mathematical induction that if $x = e^w$ the quantity $x^n \frac{d^n y}{dx^n}$ may be reduced to the form

$$a_n \frac{d^n y}{dw^n} + a_{n-1}\frac{d^{n-1}y}{dw^{n-1}} + \cdots + a_1 \frac{dy}{dw} + a_0 y,$$

where the a's are constant.

(A proof by mathematical induction consists of two steps: (1) the proof that the statement is true for one value of n; (2) the proof that if it is true for *any* value of n, it is true for the *next higher* value. It then follows that the statement is true for any value of n which can be reached by counting from the one for which it was specifically proved.)

10. An expression of the form

$$b_m x^m \frac{d^m y}{dx^m} + b_{m-1}x^{m-1}\frac{d^{m-1}y}{dx^{m-1}} + \cdots + b_1 x \frac{dy}{dx} + b_0 y$$

is called a homogeneous linear differential expression. Show that the substitution $x = e^w$ reduces it to the form

$$c_m \frac{d^m y}{dw^n} + c_{m-1}\frac{d^{m-1}y}{dw^{n-1}} + \cdots + c_1 \frac{dy}{dw} + c_0 y.$$

11. Write the equation of a family of circles all of the same radius r, the centers of which are distributed on a circle of unit radius about the origin.

12. Find the envelope of the family of curves in Problem 11.

13. Find the envelope of the family of lines

$$y = px + 2p^2.$$

14. One point of a line slides along the hyperbola $xy = 1$. Another point slides along the x-axis. What is the envelope of the lines, if the points are separated by unit distance?

15. The ends of a wire of unit length slide along two rods which are perpendicular to one another. How large an area does the wire sweep over?

16. Classify the following differential equations:

(*a*) $$\frac{d^2v}{du^2} = \sqrt[3]{\frac{1}{v} + \left(\frac{dv}{du}\right)^4}.$$

(*b*) $$\frac{dv}{du} + u^2v = \sin u.$$

(*c*) $$\frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial x^2} = \frac{\partial^2 v}{\partial t^2}.$$

17. Classify the following differential equations:

(*a*) $$\sqrt{\frac{dy}{dx} + y} = \sqrt[4]{\frac{d^2y}{dx^2} + 2x}.$$

(*b*) $$\frac{dy}{dx} + 2\frac{d^2z}{dx^2} = f(y, z).$$

18. Change the dependent variable in the equation

$$\frac{dv}{du} + \frac{2v}{u} = 3$$

to $w = ve^{2u}$.

CHAPTER II

THE SIGNIFICANCE OF DIFFERENTIAL EQUATIONS AND THEIR SOLUTIONS

§ 6. *Geometrical Interpretation of a First Order Differential Equation*

Any algebraic [1] equation of the form

$$y = f(x)$$

may be graphically represented by a curve. Since a solution of a differential equation is, by definition, a functional relation between x and y, it may also be represented by a curve. Thus, in a sense, either an algebraic equation or a differential equation defines a curve. They do not, however, use the same means to accomplish this end, and a study of their differences in this respect gives a valuable picture of the real significance of a differential equation.

In the case of an algebraic equation, when a value is chosen for x the equation defines one or more values of y which may be associated with that x. These fix one or more points on the curve. By assigning other values, other points are obtained. That is, an algebraic equation defines a curve *by relating to one another the coordinates of each point on it.*

A differential equation, on the other hand, does not relate y's to x's directly: it makes use of a totally different process. Consider, for the time being, an equation of the first order containing only the two variables x and y. Any such equation may be written in the form

$$\frac{dy}{dx} = f(x, y). \tag{19}$$

[1] Throughout this chapter the term "algebraic" is used to denote any equation which does not contain derivatives.

The substitution of a special value for x in this equation does not lead to a value for y; instead, when x and y are *both* specified a value of $\frac{dy}{dx}$ is determined. That is, if any point (x, y) is chosen, (19) determines the direction in which the curve must proceed *if it passes through that point.* Of course there may be no reason to think that the desired curve actually passes through this point. For the moment, however, we shall overlook this fact, and shall imagine that we have determined the directions corresponding to a very large number of points and denoted them by arrows, as in Fig. 9. As we look at this figure each arrow tends to carry the eye to the arrow next ahead of it, in such a way as to associate the arrows into groups. By beginning at any point A and connecting the arrows of such a group together, a curve is produced which has, at each of the points to which arrows are affixed, the slope required by (19).

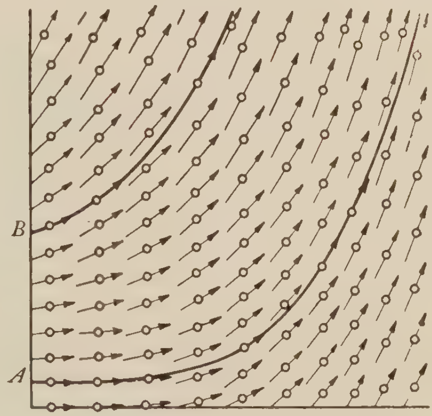

FIG. 9.

If the number of arrows is greatly increased the number of points at which the curve satisfies (19) is also greatly increased. In fact, by making the arrows more and more numerous it would be possible, in theory at least, to approach a limiting curve, the coordinates and slope of which would satisfy the differential equation at every point. This limiting curve — or rather the relation between y and x of which it is the graphical representation — is a solution of (19).

A differential equation therefore defines a curve *by telling the direction in which it passes through each of its points.*

This geometrical picture affords a graphical method of solving the differential equation. In practice it is found not

to be very accurate, but nevertheless it is occasionally used when more exact methods are too laborious. There is one obvious uncertainty about it, however, and that is the choice of the point A at which the process begins. If a different point B were chosen, a different curve would be obtained. Obviously the coordinates and slope of this new curve, like those of the old one, satisfy (19); both are therefore "solutions." *Hence, there is more than one solution of (19).*

§ 7. *The Number of Distinct Solutions of a First Order Differential Equation*

We can also obtain from this geometrical construction a valuable picture of the significance of the theorem that *there are just as many solutions of the differential equation (19) as there are points on a straight line.*

We have already noticed that one solution of (19) was obtained by starting from A and another by starting from B. Obviously we could build up other solutions by starting from other points of the line AB, and we would judge from the appearance of Fig. 9 that all these solutions would be different.

We could also build up solutions by starting from points which were *not* on the line AB, but we can readily see from the appearance of the figure that the curves thus obtained would cross AB.

In the case of Fig. 9, therefore, the theorem appears to be true; for to every point of AB there appears to correspond a curve, and there appear to be no other possibilities. We must be on our guard, however, not to be led astray by the simplicity of our diagram; and indeed we can readily find equations which lead to pictures radically different from Fig. 9, and which appear upon first thought to invalidate our theorem.

For instance, if our differential equation were

$$\frac{1}{\frac{dy}{dx}} = 0,$$

our arrows would all be vertical. Hence, upon starting from A the line AB would itself be obtained as a solution. The same solution — not a different one — would be obtained by starting from any other point of AB. But in this case *every* solution is a vertical straight line, and therefore all of them would be obtained by starting from the points of any *horizontal* line. Their number is therefore again the same as the number of points on a line.

Again the equation

$$\frac{dy}{dx} + \frac{x}{y} = 0$$

has for its solutions the circles shown in Fig. 10. It is obvious that identical solutions are obtained by starting from corresponding points on the upper and lower halves of the axis: there are therefore as many distinct solutions as there are points on *part* (not all) of the line. But a fundamental theorem in the Theory of Point-Sets states that these can be made to correspond uniquely with *all* the points on some other line, so that there are still just as many distinct solutions as there are points on this auxiliary line.[1] Hence the theorem is still valid.

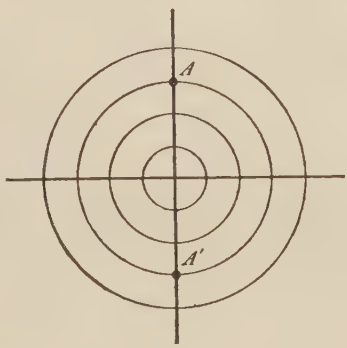

FIG. 10.

[1] This statement may appear utterly absurd. It seems dangerously near saying that half a number equals the number; and of course this latter statement is not true of numbers other than zero. The difficulty lies in the peculiar properties of the concept of infinity. Of course, no one objects to the statement that "half of infinity is still infinite."

We might, perhaps, let the matter rest there, so far as showing that the statement is not altogether absurd. Actually, however, it has a deeper meaning, which we can at least vaguely indicate by a simple example.

Every positive number has a logarithm, and conversely, to every logarithm there corresponds a positive number. There are, then, just as many positive numbers as there are logarithms. But since the logarithms of numbers less than unity are negative, we would require an entire line upon which to plot them, though their anti-

Finally, the equation

$$\frac{dy}{dx} + \frac{y}{x} = 0$$

is satisfied by the hyperbolas shown in Fig. 11. By using points on the dotted line AB, only those curves which lie in the first and third quadrants would be obtained; those in the second and fourth would be missed. But by using the dotted

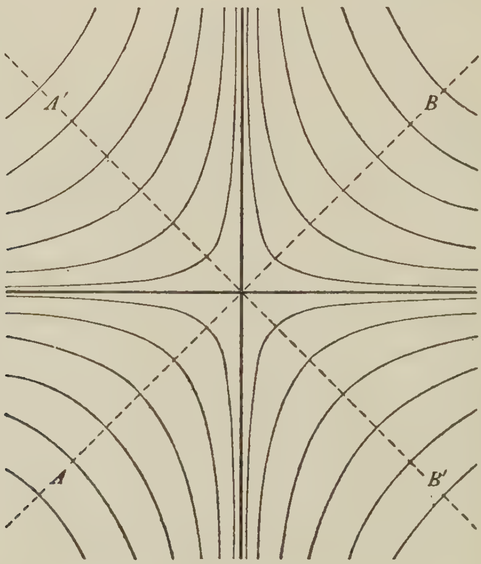

Fig. 11.

line $A'B'$ also, the latter group would be included. This would seem to require the points of *two* lines; but the same general theorem of point sets referred to above makes it possible to refer all these to the points of an auxiliary line, and the theorem remains valid.

In fact, the proof of the theorem can be made to apply as long as it is possible to divide the entire xy-plane into a countable number of regions (four in Fig. 11) in the interior of each

---

logarithms could be plotted on only half a line. Thus we have paired up the points of half a line with the points of a whole line, which is exactly the thing needed in the case of Fig. 10.

of which either the $f(x, y)$ of equation (19) or its reciprocal is finite and continuous.

Finally, we have tacitly assumed that the function $f(x, y)$ in (19) was single-valued. This would appear to rule out a large number of equations — for example, any which contained square roots. But if there were *two* values of f at every point, two arrows could be drawn through it, and therefore two curves could be obtained by starting in these two directions. It is easy to see, however, that these two sets of curves could be associated with the points of *two* auxiliary lines. But again the same point-set theorem which has already been several times used, makes it possible to say that these can be referred in turn to the two halves of a single line, and so even the restriction that f must be single-valued may be removed.

By carrying the reasoning one step further our theorem may be phrased in a more convenient form. The "solution" of (19) was made precise only by stating at what point the curve crossed the line AB. As this point is defined by its distance η above the x-axis, it follows that the exact form of solution must depend upon the value of this constant η. It must therefore be written in the form

$$\phi(x, y, \eta) = 0,$$

for this notation means nothing more nor less than that the exact relationship between x and y is known when and only when the value of η is given. As there is a value of η for every point on AB, and conversely, this relation includes every possible solution [1] of the differential equation. It is called the *general solution or primitive*, each of the relations which may be obtained from it by assigning a special value to η being designated a *particular solution*.[2] Hence the following theorem,

[1] In Chapter V we shall find that there are sometimes exceptions to this statement. That is, there are certain solutions, called "singular solutions," which do not belong to the same family as the rest. They need not concern us here, however.

[2] Note that the "general solution" corresponds to the entire family of curves which the equation defines, while the term "particular solution" refers to only one of them.

which represents the ultimate result of our present discussion, is derived:

The general solution of equation (19) is a relation between x and y containing one and only one arbitrary constant.

§ 8. *Geometrical Interpretation of a Second Order Differential Equation*

Equation (19) contains no derivatives higher than the first. Hence the geometrical construction of § 7 applies to first order equations only, and the theorem discussed in § 7 cannot be said to be true for equations of higher order. It is the purpose of the next few sections to extend the reasoning to include second order equations, from which the result for higher orders may be inferred.

Consider an equation which contains, in addition to x, y and y', the second derivative y''. Suppose as before that the highest derivative is isolated, so that the equation takes the form

$$y'' = f(x, y, y'), \tag{20}$$

f being a single-valued function. The situation is now slightly more complicated than before; but a geometrical interpretation can still be obtained, based upon the conception of the "radius of curvature" of a curve — that is, the radius of the circle which most accurately fits it. This radius of curvature ρ is known to be given by the expression

$$\rho = \frac{(1 + y'^2)^{3/2}}{y''}.$$

Hence, for any curve which is a solution of (20), the radius of curvature must be

$$\rho = \frac{(1 + y'^2)^{3/2}}{f(x, y, y')}. \tag{21}$$

Let us now choose a point A and a direction AT, Fig. 12, and require that the solution of (20) pass through A in the

direction AT. The radius of curvature ρ which the curve must have at this point is obtained at once from (21) by using for x and y the coordinates of A, and for y' the slope of the line AT. If we lay off this distance ρ on a line through A perpendicular to AT, thus locating a point C, and about this point as a center draw a short circular arc AA_1, we shall evidently produce an arc which not only passes through A

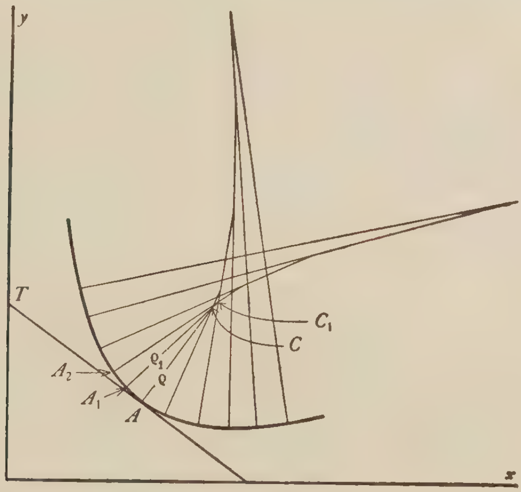

FIG. 12.

in the desired direction, but which has the curvature required by the differential equation as well.

To the end-point A_1 of this arc corresponds a new set of coordinates (x_1, y_1) and a new slope y_1', the latter being the slope of the circular arc at this point. When these new values are substituted in (21) a new radius of curvature ρ_1 is obtained. This is laid off on the line A_1C, giving the new center C_1 about which the next short arc A_1A_2 is to be described. By continuing this process, a curve $AA_1A_2 \cdot \cdot \cdot$ may be obtained which possesses the property that it leaves each lettered point in such a direction and with such a curvature that the differential equation is satisfied.

If all the arcs were made so small that the curvature did not change much from one to another, it might be expected that the curve thus drawn would differ very little from the true solution of the differential equation. In fact, from a theoretical standpoint, the correct solution could be approached as closely as desired by shortening the arcs more and more. The construction therefore gives an approximate method of solving the equation, though it is of no great use in practice since errors of draftsmanship seriously limit its accuracy.

§ 9. *The Number of Solutions Possessed by Equations of Order Higher than the First*

The importance of this geometrical interpretation of (20) lies not so much in its usefulness as a method of solution as in the information it gives regarding the family of curves defined by (20).

To start the construction of the curve $AA_1A_2 \cdots$, both the point A and the direction AT had to be chosen. If a different point B had been selected a different solution would have been obtained even though the same direction had been kept; and if a different direction AT_1 had been chosen still another solution would have been obtained, even though the same point A were used. In fact, through A pass a pencil of curves, one of which leaves it in every possible direction, and all of which are solutions of (20). The separate curves of this pencil may be distinguished by their slopes at A. Hence, if the symbol η' is used to define this slope, the entire pencil may be represented by a relationship between x and y which depends upon the arbitrary constant η'.

This pencil does not, however, contain all the solutions of (20). Instead, there is a similar pencil for every point on the vertical line through A. Hence the relationship between x and y which includes every possible solution — that is, the general solution of the differential equation — must depend on the ordinate η as well as the slope η'. This leads to the theorem:

The general solution of a differential equation of the second order is a functional relationship

$$\phi(x, y, \eta, \eta') = 0,$$

containing two and only two arbitrary constants η *and* η'.

As was the case with the discussion of an equation of the first order, many refinements are necessary if this argument is to be made rigorous. But as we seek less for a rigorous proof than for a clear picture of what our theorem means, the argument is good enough.

For the general case of an equation of the nth order a similar theorem may be built up. No attempt need be made to derive it, however, as it is fairly evident by analogy that it must take the form:

The general solution of a differential equation of the n*th order is a functional relation of the form*

$$\phi(x, y, \eta, \eta', \cdot \cdot \cdot, \eta^{(n-1)}) = 0,$$

in which the constants $\eta, \eta', \cdot \cdot \cdot, \eta^{(n-1)}$ *are arbitrary.*

§ 10. *Boundary Conditions*

To solve a differential equation, from the classroom standpoint, means to obtain its general solution. From this point of view no problem has been completely solved until a result has been derived in which the number of arbitrary constants equals the order of the equation.

From a practical standpoint the situation is somewhat different. When a differential equation arises in connection with a scientific investigation it is usually a particular solution — not the general one — which is desired; for the scientific investigation itself usually makes it obvious that the desired solution, in addition to satisfying the equation itself, must satisfy certain extraneous conditions peculiar to the investigation in hand. These are known as "boundary conditions" or "boundary values."

As an example, consider the motion of a simple pendulum, Fig. 13, held at an angle Θ, and then released. It is known that the angular acceleration $\frac{d^2\theta}{dt^2}$ of a pendulum is proportional to its angular displacement[1] θ, and that they have opposite signs. Expressed as an equation, these facts give

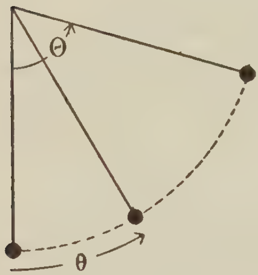

FIG. 13.

$$\frac{d^2\theta}{dt^2} = - p^2\theta, \tag{22}$$

where p is a positive constant.

By a method which will be explained later in the text, the general solution of this equation is known to be

$$\theta = \alpha \cos pt + \beta \sin pt, \tag{23}$$

α and β being the two arbitrary constants. Owing to the presence of these arbitrary constants the motion defined by (23) is quite indefinite. This may be seen by attempting to find out through how great an angle the pendulum swings. It is easy to throw (23) into the form

$$\theta = \sqrt{\alpha^2 + \beta^2} \sin (pt + \epsilon),$$

where

$$\tan \epsilon = \frac{\alpha}{\beta};$$

and in this form it is evident that the largest value ever attained by θ is $\sqrt{\alpha^2 + \beta^2}$. But since α and β are arbitrary, this maximum value may be anything at all.

On the other hand, there is nothing indefinite about the physical motion of the pendulum. It is a matter of common experience that if it is held quietly at the angle Θ and then released, its future history is completely determined. It follows, of course, that *some one set* of values of α and β must

[1] This is true to a high degree of approximation where θ is small, provided there are no dissipative forces, such as the friction of bearings or the resistance of the air.

give the proper solution, the other sets corresponding to the possible histories of the pendulum if released in other ways. Hence, from the practical standpoint, the problem can only be said to be completely solved when the appropriate α and β have been found.

In the case at hand this is easy enough. If the instant when the pendulum is released is called $t = 0$, it is known: first, that at that time θ was equal to Θ, and second, that the velocity $\frac{d\theta}{dt}$ was zero. Expressed in the form of equations these conditions are

$$\left.\begin{aligned} \theta &= \Theta \quad \text{when } t = 0, \\ \frac{d\theta}{dt} &= 0 \quad \text{when } t = 0. \end{aligned}\right\} \tag{24}$$

But from (23)

$$\left.\begin{aligned} \theta &= \alpha \quad \text{when } t = 0, \\ \frac{d\theta}{dt} &= p\beta \quad \text{when } t = 0. \end{aligned}\right\} \tag{25}$$

Comparing (24) and (25) it is obvious that we must make $\alpha = \Theta$ and $\beta = 0$. Hence the desired solution is

$$\theta = \Theta \cos pt.$$

If θ and t are plotted as ordinate and abscissa respectively in a Cartesian graph, these boundary conditions (24) tell *where* and *in what* direction the curve crosses the vertical axis. It is interesting to note that this is exactly the information which had to be assumed in building up the geometrical picture of § 9. There are other types of conditions which might be used, however, and which would serve the purpose equally well. For instance, with the pendulum already swinging, its position might be noted at two instants a second apart. If these were called $t = 0$ and $t = 1$, and if the corresponding positions were $\theta = \theta_0$ and $\theta = \theta_1$, (23) would give immediately

$$\theta_0 = \alpha,$$

$$\theta_1 = \alpha \cos p + \beta \sin p.$$

The constants α and β would then be

$$\alpha = \theta_0,$$
$$\beta = \theta_1 \csc p - \theta_0 \cot p;$$

and the desired solution

$$\theta = \theta_0 \cos pt + (\theta_1 \csc p - \theta_0 \cot p) \sin pt.$$

This may be reduced to the somewhat simpler form

$$\theta = \frac{\theta_0 \sin p(1 - t) + \theta_1 \sin pt}{\sin p}.$$

Graphically, this set of boundary conditions fixes, not a point and the direction in which the curve passes through that point, but two points on the curve instead, leaving the directions to take care of themselves. Many other types of boundary conditions might work equally well; for example, the angular velocities of the pendulum as observed at two different instants, but not the positions; or the position at one instant and the velocity at another, and so on. Formulated as equations these would be

$$\left.\begin{aligned} \frac{d\theta}{dt} &= \theta'_0 \quad \text{at} \quad t = t_0, \\ \frac{d\theta}{dt} &= \theta'_1 \quad \text{at} \quad t = t_1, \end{aligned}\right\} \tag{26}$$

and

$$\left.\begin{aligned} \theta &= \theta_0 \quad \text{at} \quad t = t_0, \\ \frac{d\theta}{dt} &= \theta'_1 \quad \text{at} \quad t = t_1, \end{aligned}\right\} \tag{27}$$

respectively. Graphically, the first would give the slope of the curve where it intersected each of two vertical lines, without specifying the ordinates of the points of intersection, while the second would give one point through which it passed and its slope where it crossed some other ordinate.

It will be noted that in this example the process of solution has consisted of two parts: first, the determination of the general solution; then the choice of the constants in such a way as to satisfy the boundary conditions. This is usually, though not quite always, the route by which a boundary value problem is solved. It has already been said that, from the classroom standpoint the equation is said to be solved as soon as the first of these steps has been taken, and it might seem at first sight that this attitude ignored a very important phase of the subject. But it must be noted that, once the general solution has been obtained, the determination of the arbitrary constants requires only the solution of a set of equations, none of which contains derivatives. Logically, therefore, that part of the problem belongs to another branch of mathematics — the Theory of Equations — and not to the study of Differential Equations. The classroom attitude is therefore not entirely without justification.

§ 11. *The Existence Proof*

In § 10 boundary conditions were stated in several forms, and in each case a solution of (22) was found which satisfied these conditions. It is not unnatural to ask whether a solution can always be found, no matter in what form the boundary values are given. That this question must be answered in the negative can be shown by presenting a case which has no solution.

Suppose it is desired to find a solution of (22) satisfying the conditions:

$$\left.\begin{aligned} \theta &= \theta_0 \quad \text{when } t = 0, \\ \theta &= 2\theta_0 \text{ when } t = \frac{\pi}{p}. \end{aligned}\right\} \qquad (28)$$

When these conditions are substituted in (23) they give

$$\theta_0 = \alpha,$$

$$2\theta_0 = -\alpha.$$

Obviously α cannot satisfy both of these equations; therefore no such solution exists. The problem set is an impossible one, and any attempt to solve it is bound to fail.

There is a physical explanation of why this set of conditions cannot be satisfied. The swing of the pendulum in one direction is a perfect copy of its swing in the other, except that at corresponding moments it is on opposite sides of the vertical, and is traveling in opposite directions. Therefore, if θ is θ_0 at a certain instant, it must be $-\theta_0$ after the lapse of a time equal to that required for a complete swing. To require it to be somewhere else, as is done by (28), is physically absurd.

This answers negatively the question as to whether all boundary conditions can be met, but it gives no way of telling in advance whether a particular set of conditions is possible or impossible. Usually, when the differential equation arises in connection with some scientific problem, the problem in question is known to have an answer, and therefore it is a mere matter of common sense that its mathematical formulation must also have an answer. But such an argument is indirect at best, and does not give criteria that have general usefulness. Sometimes, however, scientific problems deal directly with the question of whether a given state is possible or not, and then this common-sense argument is of no use at all. Even at its best it is indirect, and is always unsatisfactory when dealing with purely mathematical questions. Hence mathematicians have given a considerable amount of attention to "existence proofs": that is, to the determination of those boundary conditions for which it can be affirmed that a solution exists. It is quite beyond the scope of this text to present the methods by means of which such studies are made, or even to give any considerable part of their results. One particular result, however, bears such a close analogy to the geometrical picture which we have just presented that it is worth brief consideration.

This result deals with the case in which the boundary conditions are given by stating values of y and its derivatives at

one and the same value of x: that is, the boundary values are

$$\left.\begin{aligned} y &= \eta, \\ y' &= \eta', \\ &\vdots \\ y^{(n-1)} &= \eta^{(n-1)}, \end{aligned}\right\} \quad \text{all for } x = \xi. \tag{29}$$

It was exactly this type of boundary values which was used in §§ 6 to 9. If, as in those sections, the differential equation is assumed to have been solved for its highest derivative, so that it appears in the form

$$y^{(n)} = f(x, y, y', \cdots, y^{(n-1)}),$$

it is found always to have a solution satisfying the boundary conditions (29), provided the function f satisfies two conditions[1]: (1) It is continuous at $x = \xi$, $y = \eta, \cdots, y^{(n-1)} = \eta^{(n-1)}$; (2) It can be differentiated with respect to $y, y', \cdots, y^{(n-1)}$. The meaning of these conditions has been explained in §§ 3 and 4.

PROBLEMS

1. Determine the arbitrary constants in (23) so as to satisfy the conditions (26).

2. Determine the arbitrary constants in (23) so as to satisfy (27).

---

[1] It is also implied that the function f is single-valued. If f is multiple-valued there will be as many solutions satisfying the conditions (29) as there are distinct values of f.

CHAPTER III

The Origin of Differential Equations

§ 12. *Origin of Differential Equations*

Before beginning to explain how differential equations are solved, a few words may be appropriate as to how they originate. In the present chapter we shall discuss two different ways. The discussion of the first of these is really the converse of the argument presented in Chapter II, and like that argument leads to the rule that the number of arbitrary constants in the solution of a differential equation is equal to the order of the equation. It is important from a theoretical, rather than a practical, standpoint; for the differential equations with which a student is actually confronted in practice seldom originate in this first fashion. Instead they generally present themselves immediately as a formulation of physical laws in terms of differential symbols. The second portion of the chapter is devoted to a few illustrations of this latter sort.

§ 13. *Development of a Differential Equation from Its Primitive*

We have seen that the general solution of a differential equation is representable as a *family* of curves. Therefore it is proper to call the equation *the differential equation of the family*. It is the purpose of the present section to explain how, when a family of curves is given, its differential equation may be obtained.

As we know, a family of curves is defined by an equation

$$f(x, y, c) = 0, \tag{30}$$

which contains, besides the variables x and y, an arbitrary constant c. The different curves of the family correspond to

different values of c. By actual differentiation a new equation

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0 \tag{31}$$

may be obtained. If this equation does not involve c, it is the differential equation desired. If c does appear in it, as is usually the case, the pair of equations (30) and (31) may be used to eliminate c, thus giving rise to a new equation involving only x, y and $\frac{dy}{dx}$. This will then be the differential equation of the family of curves (30).

For instance, suppose the family is

$$y = \sin cx. \tag{32}$$

Differentiation gives

$$y' = c \cos cx. \tag{33}$$

Both of these equations involve c and hence neither is the desired equation. But from (32) we find that

$$\cos cx = \sqrt{1 - y^2},$$

$$c = \frac{1}{x}\sin^{-1} y;$$

and when these values are substituted in (33) they give

$$\frac{dy}{dx} = \frac{1}{x}\sqrt{1 - y^2}\sin^{-1} y. \tag{34}$$

As this is entirely independent of c it is the differential equation of the family (32).

This process of eliminating arbitrary constants can be carried out even when the number of independent constants is greater than 1. To show this, let

$$f(x, y, c_1, c_2, \cdots, c_n) = 0 \tag{35}$$

be an equation connecting x and y and involving the n constants $c_1, c_2, \cdots, c_n$. This equation also represents a family of curves, and the problem before us is that of deriving a dif-

ferential equation which will define the set without the use of arbitrary constants. If (35) is differentiated n times in succession with respect to x a set of n other equations is obtained. These equations, together with (35) itself, constitute the set of $n + 1$ equations necessary to eliminate the n c's. The result of this elimination is the desired differential equation. As none of the equations involves derivatives of higher order than the nth, the resulting differential equation cannot be of order higher than n. Conversely, the order of the equation cannot be less than n, unless the nth derivative of (35) was not required in the solution, in which case the constants were not all independent. Hence: *the order of the highest derivative in the differential equation is the same as the number of independent constants in the primitive.* This is the rule already obtained in Chapter II for the number of constants in the solution of a differential equation.

As an example, consider the equation

$$\theta = \alpha \sin pt + \beta \cos pt \tag{23}$$

containing the variables θ and t and the arbitrary constants α and β. By differentiating twice we get

$$\theta' = \alpha p \cos pt - \beta p \sin pt,$$

$$\theta'' = - \alpha p^2 \sin pt - \beta p^2 \cos pt.$$

In this case elimination is very simple since inspection gives immediately

$$\frac{d^2\theta}{dt^2} + p^2\theta = 0. \tag{22}$$

This is the desired differential equation. It has already been met in § 10, where (23) was arbitrarily stated to be its solution.

§ 14. *Development of Differential Equations from Physical Laws*

In the application of mathematics to engineering and physical sciences differential equations occur repeatedly, for many physical laws can be more simply expressed by means of differential symbols than in any other way. An illustration

has already appeared in § 10, where the physical laws governing the motion of a simple pendulum gave rise at once to a differential equation of the second order. In the sections which follow several other examples will be given.

Some of these examples require the use of physical laws with which the student may not be familiar. In each case, therefore, a brief statement of the scientific facts necessary for the understanding of the problem is given. No attempt is made to explain how these facts are known to be true — to do that would require the teaching of the branches of physics from which the illustrations are drawn, which is obviously beyond the scope of the book. From our standpoint the laws are needed for only one purpose: to show that the differential equations follow naturally and directly from them — so directly, in fact, that were the student familiar with the subjects with which they deal he would find little difficulty in formulating the same equations for himself.

This being the purpose of the discussions, the student will be wise to be content when he is sure he understands the meaning of the scientific statements, and not to speculate too much upon how their truth may have been established. In other words, he will succeed best if he approaches the preliminary statements with a judicious amount of faith, reserving his usual healthy skepticism for the mathematical arguments by which they are followed.

§ 15. *Example 1. The Law of Mass Action*

Chemical reactions do not take place instantaneously when the reagents[1] are brought in contact. There is a certain amount of compound formed in the first thousandth of a second, some more in the second thousandth, and so on. The rate of formation of the compound is not constant, however, but varies with the amounts of the reagents which are available for forming the compound. In many simple transformations the speed of reaction obeys a law known as the "law of mass action," which states that the rate at which the simple sub-

[1] The "reagents" are the active chemicals.

stances are converted into the compound is proportional to the product of the amounts of the substances which are still unconverted.

Suppose that x is the number of compound molecules already formed at the time t, and that each compound molecule contains n molecules of one kind and m of another. Finally, assume that to start with there were N molecules of the first kind and M of the second. Then the number of unconverted molecules of each kind is $N - nx$ and $M - mx$, respectively. The law of mass action states that the rate at which new molecules of the compound are formed—which is obviously given by $\frac{dx}{dt}$ — is proportional to the product of these last two quantities. This gives immediately the differential equation

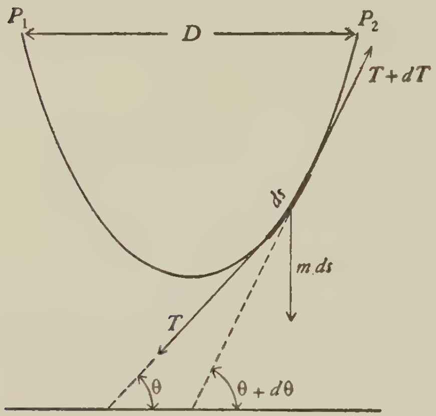

Fig. 14.

$$\frac{dx}{dt} = k(N - nx)(M - mx). \tag{36}$$

§ 16. *Example 2. The Curve of Suspension of a Flexible Inextensible Cord*

A cord is suspended from two points P_1 and P_2 on the same level and at a distance D from one another. It is desired to find the shape assumed by this cord under the action of gravity.

Let the curve of Fig. 14 represent this shape, and consider any element of length ds.[1] It is one of the fundamental proposi-

[1] Properly, of course, this element is an "increment in length" and would be denoted in the Calculus by Δs. It is customary in scientific work, however, when such an increment is eventually to be made to vanish, to use the differential symbol throughout. The practice is seldom confusing, and often serves the useful purpose of forecasting what the nature of the argument is to be.

tions in mechanics that this element must be in equilibrium under the forces acting upon it. These forces are

(*a*) Its own weight, which is a force acting vertically downward;

(*b*) The tension of the rope at the lower end, which acts in the direction of the tangent at that point;

(*c*) The tension in the rope at the upper end, which acts in the direction of the tangent to the curve at that point.

Call the inclination of the tangents at the two ends θ and $\theta + d\theta$ and the tensions T and $T + dT$, and denote the linear weight [1] of the cord by m. Then if the three forces are resolved into their X and Y components, they give, respectively,

$$X_1 = 0, \qquad Y_1 = -m\,ds,$$

$$X_2 = -T\cos\theta, \qquad Y_2 = -T\sin\theta,$$

$$X_3 = (T + dT)\cos(\theta + d\theta), \qquad Y_3 = (T + dT)\sin(\theta + d\theta).$$

If the element of the cord is to be in equilibrium under the action of these forces it is necessary that both the sum of the X components and the sum of the Y components shall be zero. That is

$$\left.\begin{aligned} T\cos\theta &= (T + dT)\cos(\theta + d\theta), \\ T\sin\theta &= (T + dT)\sin(\theta + d\theta) - m\,ds. \end{aligned}\right\} \qquad (37)$$

Dividing these equations member by member,

$$\tan\theta = \tan(\theta + d\theta) - \frac{m\,ds}{(T + dT)\cos(\theta + d\theta)}. \qquad (38)$$

The first of equations (37) states that the *horizontal* component of the tension is the same at both ends of the element ds. As the element ds is any element whatsoever it follows that this component is the same at every point of the curve.[2] If it is

[1] That is, the weight per unit length.

[2] It is actually the *least* tension at any point of the curve; that is, the tension at the bottom where the vertical component of tension vanishes.

called κ, (38) becomes

$$\tan\theta = \tan(\theta + d\theta) - \frac{m\,ds}{\kappa}.$$

If this is written in the form

$$\frac{\tan(\theta + d\theta) - \tan\theta}{d\theta} = \frac{m}{\kappa}\frac{ds}{d\theta}$$

and ds and $d\theta$ are allowed to vanish, the left-hand side of the equation degenerates into the derivative of $\tan\theta$. Hence

$$\sec^2\theta = \frac{m}{\kappa}\frac{ds}{d\theta}. \tag{39}$$

This is the differential equation of the desired curve in what mathematicians call its *intrinsic form*; that is, it expresses the length s measured from some arbitrary point in terms of θ, the inclination of the tangent. For many purposes, however, the intrinsic form is not very serviceable and it is therefore best to reduce it to the ordinary Cartesian form. This requires a change of *both* of the variables s and θ to new variables x and y related to them by the laws

$$\frac{dy}{dx} = \tan\theta, \tag{40}$$

$$\left(\frac{ds}{dx}\right)^2 = 1 + \left(\frac{dy}{dx}\right)^2.$$

The variable s is eliminated by noting that

$$\frac{ds}{d\theta} = \frac{\dfrac{ds}{dx}}{\dfrac{d\theta}{dx}}.$$

This gives

$$\sec^2\theta\,\frac{d\theta}{dx} = \frac{m}{\kappa}\frac{ds}{dx} = \frac{m}{\kappa}\sqrt{1 + \left(\frac{dy}{dx}\right)^2}.$$

Also by differentiation of (40) it is found that

$$\sec^2\theta\,\frac{d\theta}{dx} = \frac{d^2y}{dx^2}.$$

The result of the substitutions is therefore

$$\frac{d^2y}{dx^2} = \frac{m}{\kappa}\sqrt{1 + \left(\frac{dy}{dx}\right)^2}. \tag{41}$$

This is the differential equation of the curve assumed by the cord, expressed in terms of the Cartesian coordinates x and y.

§ 17. *Example 3. The Flow of Current in an Electrical Network*

Five laws governing the flow of electricity in a network containing resistances, inductances and capacities are:

(i) The amount which flows into any junction-point, such as b, Fig. 15, must be equal to the amount which flows out.

(ii) The algebraic sum of all the electromotive forces met in passing around a complete loop, such as *abeda* or *bcfeb*, must vanish.

(iii) The electromotive force across an element of resistance R is equal to the product of R by the current flowing through it.

(iv) The electromotive force across an element of inductance L is equal to the product of L by the rate at which the current is changing.

(v) The electromotive force across an element of capacity C is equal to the charge on the condenser divided by C.

In order to formulate these statements in mathematical form, let x denote the quantity of electricity which has passed a given point since the time $t = 0$; x_0 the quantity upon the condenser at $t = 0$; R, L and C the magnitudes of the resistance, inductance and capacity, and I the current[1] flowing. Then I is related to x by the equation

$$I = \frac{dx}{dt} \quad \text{or} \quad x = \int_0^t I\,dt,$$

and the electromotive forces E_R, E_L and E_C across a resistance,

[1] "Current" is the rate at which electricity is transported.

an inductance and a capacity, respectively, are given by the formulæ

$$E_R = RI = R\frac{dx}{dt},$$

$$E_L = L\frac{dI}{dt} = L\frac{d^2x}{dt^2},$$

$$E_C = \frac{x + x_0}{C} = \frac{x_0}{C} + \frac{1}{C}\int_0^t I\,dt.$$

By means of these formulæ and the use of the laws (i) and (ii) it is possible to write down at once the differential equations which must be obeyed by either the quantity x or the current I.

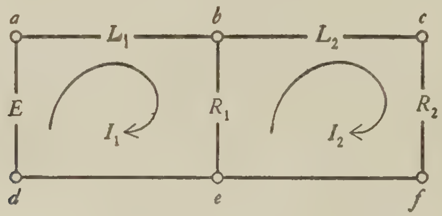

FIG. 15.

When rule (i) is applied to the circuit of Fig. 15 it leads to the conclusion that the same current I_1 flows through the branches *ed*, *da* and *ab*; and the same current I_2 through the branches *bc*, *cf* and *fe*, while the current through the one remaining branch *be* is[1] $I_1 - I_2$.

Then the application of rules (iii) and (iv) gives the following electromotive forces:

in *ab*,	$L_1\frac{dI_1}{dt}$;
in *bc*,	$L_2\frac{dI_2}{dt}$;
in *cf*,	R_2I_2;
in *be*,	$R_1(I_1 - I_2)$;

[1] I_2 is subtracted because it flows in the opposite direction from I_1.

while in *fe* and *ed* the electromotive force is zero since there are no impedance elements there. In *da* there is supposed to be a generator producing the electromotive force E which gives rise to the current flow.

Having these values, the application of rule (ii) to the meshes *abeda* and *bcfeb* separately gives the desired differential equations:

$$\left.\begin{aligned} L_1\frac{dI_1}{dt} + R_1(I_1 - I_2) + E = 0, \\ L_2\frac{dI_2}{dt} + R_2I_2 - R_1(I_1 - I_2) = 0. \end{aligned}\right\} \qquad (42)$$

The only treacherous point about the formulation of these equations lies in determining the signs of the various terms, particularly in the case of the branch *be*. This, of course, is a technical matter about which the electrical engineer would be well informed. For the present it is sufficient to note that we always sum electromotive forces by starting from a point such as *a*, and passing continuously around a loop such as *abeda*, back to *a*. Whenever the arrow which represents a current is drawn in the direction in which we are going, as in the case of I_1, the electromotive forces to which it gives rise have the positive sign. But if the arrow points against us, as in the case of I_2 when passing from b toward e, the electromotive force enters with a negative sign.

This example differs from the preceding ones in that, though there is only one independent variable t, there are two dependent variables I_1 and I_2. There are, however, two equations from which to determine them. In a later section it will be found that their solution is not difficult.

§ 18. *Example 4. The Conduction of Heat*

The temperature at every point of a body may be represented by a function $\theta(x, y, z)$. If the temperature is the same at all points θ is a constant, otherwise it varies with one or more of the coordinates. If a line AB of length Δs is drawn inside

the body and the temperatures at its end-points are observed to be θ_A and θ_B, the ratio

$$\frac{\theta_B - \theta_A}{\Delta s} = \frac{\Delta\theta}{\Delta s}$$

measures the average rate of change of θ between A and B. The limit of this ratio, as Δs is allowed to vanish, is called the *gradient* of the temperature in the direction [1] s, and is denoted by the differential symbol $\frac{\partial\theta}{\partial s}$. It signifies the rate at which θ would be observed to change by an observer who traveled away from A in the direction s. When the direction s happens to coincide with one of the coordinate axes, this gradient becomes a partial derivative of the usual sort.

With this definition understood, three fundamental laws upon which the study of heat conduction rests are:

(i) The amount of heat which a body contains per unit volume is proportional to its temperature.

(ii) The quantity of heat transported in unit time across any plane area inside the body is proportional to the product of the area by the temperature gradient normal to that area.

(iii) The heat flows from the high temperature to the low temperature side.

The constants of proportionality in (i) and (ii) are called the *thermal capacity* and the *thermal conductivity* of the body, respectively, and will be denoted by c and k.

Consider an elementary cube with the edges dx, dy and dz, one corner of which is supposed to be at the point (x, y, z). Consider first the two faces perpendicular to the axis of x. The rate at which heat is conveyed *out* of the cube across the nearer one of these faces is, by (ii) and (iii),

$$k\left.\frac{\partial\theta}{\partial x}\right|_{x,\,y,\,z} dy\,dz\,;$$

[1] "Direction s" means, of course, the direction in which the line AB was drawn.

and the rate at which it is conveyed *in* across the farther one is

$$k \frac{\partial \theta}{\partial x}\bigg|_{x+dx,\, y,\, z} dy\, dz.$$

The difference between these is the net rate at which heat is gained, so far as these two faces are concerned. If the symbol q_x is used to represent the net amount which has flowed into the cube across these two faces since the time $t = 0$, the net rate of gain must be written as $\frac{dq_x}{dt}$. Hence

$$\frac{dq_x}{dt} = k \left(\frac{\partial \theta}{\partial x}\bigg|_{x+dx,\, y,\, z} - \frac{\partial \theta}{\partial x}\bigg|_{x,\, y,\, z} \right) dy\, dz.$$

If dx is sufficiently small the bracketed quantity is approximately equal to $\frac{\partial^2 \theta}{\partial x^2}\, dx$, the approximation becoming exact if dx is allowed to vanish. Hence we have

$$\frac{dq_x}{dt} = k \frac{\partial^2 \theta}{\partial x^2}\, dx\, dy\, dz.$$

A similar argument may be applied to the faces perpendicular to the y-axis, and leads to the result that the net gain of heat through these faces takes place at the rate

$$\frac{dq_y}{dt} = k \frac{\partial^2 \theta}{\partial y^2}\, dx\, dy\, dz.$$

Through the remaining pair of faces the net gain is at the rate

$$\frac{dq_z}{dt} = k \frac{\partial^2 \theta}{\partial z^2}\, dx\, dy\, dz.$$

The total rate of gain in heat, $\frac{dq}{dt}$, is the sum of these three terms. As it takes place in a volume $dv = dx\, dy\, dz$, the net rate of gain *per unit volume* is

$$\frac{1}{dv} \frac{dq}{dt} = k \left(\frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial y^2} + \frac{\partial^2 \theta}{\partial z^2} \right). \tag{43}$$

All this comes about through the application of the second

and third laws. We are now ready to apply the first. The quantity of heat per unit volume is $c\theta(t)$ at the time t, and $c\theta(t + dt)$ at the time $t + dt$. The net gain is therefore

$$c[\theta(t + dt) - \theta(t)],$$

and as this takes place during the interval dt, the rate of gain is

$$c\,\frac{\theta(t + dt) - \theta(t)}{dt} = c\,\frac{\partial\theta}{\partial t}. \tag{44}$$

The heat thus acquired must, of course, come across the boundaries of the element (unless there are sources of heat in the element itself, which is not contemplated in this discussion), wherefore it follows that (43) and (44) must be equal. This leads to the differential equation

$$\frac{\partial^2\theta}{\partial x^2} + \frac{\partial^2\theta}{\partial y^2} + \frac{\partial^2\theta}{\partial z^2} = \frac{c}{k}\,\frac{\partial\theta}{\partial t}. \tag{45}$$

This equation must be satisfied by the temperature of any body through which heat is being conducted. It is of different form from any so far met in that, though there is only one dependent variable θ, there are four independent variables, x, y, z and t. The derivatives are therefore *partial* derivatives, and the equation is a *partial differential equation*.

As a simple example, consider a large plane slab, the faces of which have been maintained at the temperatures θ_1 and θ_2 so long that the temperature of every point of the interior has already reached a state of equilibrium and is no longer changing. Then $\frac{\partial\theta}{\partial t}$ is zero. Conditions of symmetry show that the temperature will be the same at all points of any plane parallel to the faces, which means, of course, that $\frac{\partial^2\theta}{\partial y^2}$ and $\frac{\partial^2\theta}{\partial z^2}$ are zero. Hence (45) reduces to

$$\frac{d^2\theta}{dx^2} = 0,$$

there being no longer any reason for using the round ∂'s since only one independent variable remains.

This equation is so simple that it may be solved immediately, without waiting for the development of elaborate methods. Simple integration gives

$$\theta = ax + b, \tag{46}$$

a and b being the constants of integration.

If the coordinates of the faces are $x = 0$ and $x = X$, respectively, the boundary conditions are

$$\theta_1 = b,$$

$$\theta_2 = aX + b.$$

From these, a and b are found to be

$$b = \theta_1,$$

$$a = \frac{\theta_2 - \theta_1}{X};$$

wherefore (46) becomes

$$\theta = \theta_1 + \frac{x}{X}(\theta_2 - \theta_1).$$

The graph of θ against x is therefore a straight line, having the ordinate θ_1 at the face $x = 0$ and the ordinate θ_2 at the face $x = X$.

§ 19. *Example 5. Irrotational Motion in a Perfect Fluid*

When a fluid is in motion its various elements may have quite different velocities. If only the x-component of motion is considered, it is obvious that its value at each point is a function of the coordinates x, y and z, and possibly also of the time t. This function might be denoted by $u(x, y, z, t)$. Similarly, the y-component of motion may be a function $v(x, y, z, t)$ of these same variables; and the z-component a function $w(x, y, z, t)$.

Now it can be shown — though it is not feasible to do so here — that these three functions are very often [1] equal to the three partial derivatives of one and the same function $\phi(x, y, z, t)$

[1] Specifically, when the fluid is "perfect" and its motion is "irrotational."

with respect to x, y and z. This function is known as the *velocity potential* of the motion. Moreover, the partial derivative of this same function ϕ with respect to t is proportional[1] to the *condensation* at (x, y, z); that is, to the percentage by which the instantaneous density at that point exceeds the average density of the fluid. Written as equations, these statements become

$$u = \frac{\partial \phi}{\partial x},$$

$$v = \frac{\partial \phi}{\partial y},$$

$$w = \frac{\partial \phi}{\partial z},$$

$$s = -c^2 \frac{\partial \phi}{\partial t}.$$

Assuming these statements to be true, the differential equation which the motion of the fluid must satisfy may be set up as follows:

Consider the element of volume shown in Fig. 16. A mass of fluid $\rho u\, dy\, dz\, dt$ flows into this element across the left-hand perpendicular face in time dt; and a quantity $\rho\left(u + \frac{\partial u}{\partial x} dx\right) dy\, dz\, dt$ flows out at the opposite face. Likewise a quantity $\rho v\, dx\, dz\, dt$ flows in through the front face, and a quantity $\rho\left(v + \frac{\partial v}{\partial y} dy\right) dx\, dz\, dt$ flows out at the back. Across the bottom face $\rho w\, dx\, dy\, dt$ flows in, and at the top $\rho\left(w + \frac{\partial w}{\partial z} dz\right) dx\, dy\, dt$ flows out. Taking account of all of

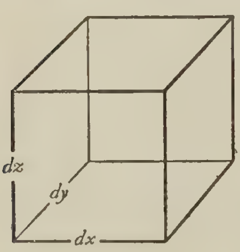

FIG. 16.

[1]The constant of proportionality is always negative. This accounts for the form in which the fourth equation below is written.

We note, also, that by definition $s = (\rho - \rho_0)/\rho_0$, s being the condensation and ρ the density. Hence we have the relation $\rho = \rho_0(s + 1)$, which we shall use later in our argument.

these and assuming that ρ differs but a negligible amount from ρ_0 throughout the element of volume, it is seen that the quantity flowing *out* exceeds the quantity flowing *in* by the amount

$$\rho_0\, dx\, dy\, dz\, dt \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right);$$

or, if it is more convenient to state it so, the quantity within the element has been increased by the negative of this amount.

This must be accounted for by a change ds in the condensation. However, the original mass of the element was $\rho_0(1 + s)\, dx\, dy\, dz$, and its final mass $\rho_0\left(1 + s + \frac{\partial s}{\partial t}\, dt\right) dx\, dy\, dz$. Therefore, the increase is $\rho_0 \frac{\partial s}{\partial t}\, dx\, dy\, dz\, dt$. Equating these two quantities, we obtain

$$\frac{\partial s}{\partial t} + \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0,$$

or

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = c^2 \frac{\partial^2 \phi}{\partial t^2}. \tag{47}$$

This is the desired equation.

If the fluid is incompressible, s must be zero everywhere and at all times. In this case (47) reduces to

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = 0. \tag{48}$$

The three equations (45), (47) and (48) are known as "the equation of heat conduction," "the wave equation" and "Laplace's equation," respectively. They are among the most fundamental equations of applied mathematics, and as they all contain the same combination of derivatives,

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2},$$

a special symbol ∇^2 has been invented to represent it. Using

this symbol, which is called "nabla squared," (45), (47) and (48) can be rewritten in the forms

$$\nabla^2\theta = \frac{c}{k}\frac{\partial\theta}{\partial t}, \tag{45}$$

$$\nabla^2\phi = c^2\frac{\partial^2\phi}{\partial t^2}, \tag{47}$$

$$\nabla^2\phi = 0. \tag{48}$$

There is a fourth differential equation of immense importance in applied mathematics which involves the same combination of derivatives as (45), (47) and (48). It is

$$\nabla^2\phi + 4\pi\rho = 0 \tag{49}$$

in which ϕ is electrical potential and ρ is the density of electricity, that is, the quantity of electricity per unit volume. This equation, which is known as "Poisson's equation," represents the distribution of potential in a region throughout which charges of electricity are scattered about in any arbitrary manner. The ρ, of course, is a function of x, y and z which tells where the charge is located. In case no charge at all is present ρ is zero, and (49) reduces to Laplace's equation.

§ 20. *Example 6. The Equation of the Potential Distribution in a Vacuum Tube*

As a last example, let us consider the potential distribution between a pair of parallel plane electrodes in a vacuum tube, one of which is supposed to be so hot that it emits electrons which are drawn across to the other by a potential difference maintained between them. According to the statements made in the last paragraph, the case is covered by the differential equation (49), once the distribution of electricity in the space between the electrodes is known. The real problem, therefore, is to find this distribution.

Now it is a fundamental physical law, that when a negative electric charge of magnitude e is transported from a place at which the potential is ϕ_1 to a place at which the potential is ϕ_2, an amount of work $(\phi_2 - \phi_1)e$ is done upon it, and its kinetic

energy is therefore increased by this amount. Let us suppose, then, that the curve of Fig. 17 represents the desired potential distribution, and assume that n electrons per second are emitted from a square centimeter of A, that each carries a charge e and has an initial velocity v_0 and a mass m. When such an electron has reached the point P at which the potential is ϕ, its kinetic energy will have been increased by an amount $(\phi - V_0)e$. Since its initial kinetic energy was $\frac{1}{2}mv_0^2$ its kinetic energy at P must be $\frac{1}{2}mv_0^2 + (\phi - V_0)e$. This, however, is equal to $\frac{1}{2}mv^2$, v being its velocity at P. Hence

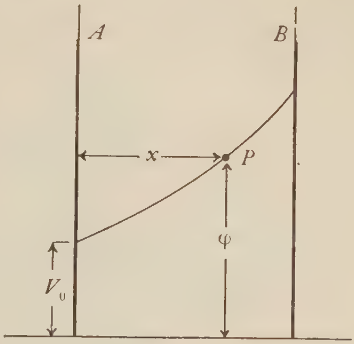

FIG. 17.

$$v = \sqrt{v_0^2 + \frac{2e}{m}(\phi - V_0)}.$$

From the number of electrons passing P per second and the velocity with which they travel, it is easy to find the space charge ρ. For this purpose attention is directed to the electrons which pass through a square centimeter of area normal to the electron stream during a short test interval dt. Those which crossed the surface at the beginning of this test interval will have traveled a distance $v\,dt$ before it ends, while those which crossed at the end of the interval will scarcely have moved at all. Hence the group under observation will be spread out over a parallelopiped of unit cross-section and height $v\,dt$. Their number is obviously $n\,dt$, and their aggregate charge $-\,ne\,dt$. But if this negative charge $-\,ne\,dt$ is distributed through the volume $v\,dt$ of our parallelopiped, its density must be

$$\rho = -\frac{ne}{v}.$$

Hence (49) becomes

$$\nabla^2\phi = \frac{4\pi ne}{v}.$$

Furthermore, because of the symmetry of the problem the potential can vary in the direction of x only and therefore $\frac{\partial^2\phi}{\partial y^2}$ and $\frac{\partial^2\phi}{\partial z^2}$ must be zero. Hence (49) becomes

$$\frac{d^2\phi}{dx^2} = \frac{4\pi ne}{\sqrt{v_0^2 + \frac{2e}{m}(\phi - V_0)}}. \tag{50}$$

This differential equation expresses the law of variation of potential inside a thermionic vacuum tube with plane electrodes, provided all the electrons are emitted with the same velocity. It is the equation from which the so-called "three-halves power law" is derived.

PROBLEMS

1. Classify all the differential equations of Chapters I and II as to kind (ordinary or partial), degree and order, and state whether they are linear or non-linear, and if linear whether the coefficients are variable or constant.

2. In the case of each ordinary equation, state the number of independent constants which will appear in the general solution.

3. Which of the solutions obtained in the examples have been general solutions? Write down two particular solutions of each example for which a general solution has been obtained. How many more are there?

4. To a fair order of approximation it may be said that the quantity of heat lost per second by a warm body is proportional to the difference in temperature between the body and the surrounding medium. Assuming this rule, write the differential equation which expresses the reading of a thermometer as a function of the time, when the bulb is transferred from a hot to a cold medium.

5. The curvature of a certain curve is at every point proportional to the slope of the normal drawn through that point. What is the differential equation of the curve?

6. Find the differential equation of the family of circles

$$(x - c)^2 + y^2 = r^2. \tag{18}$$

7. Find the differential equation of the family of curves

$$(y - c)^2 = (x + c)^3. \tag{14}$$

8. Find the differential equation of the family of curves

$$y^2 = (x - c)(x - 2c)^2. \tag{17}$$

9. A boy, standing upon the corner A of a square field, holds in his hand one end of a rope of length L. This rope is lying at full length along the side AB of the field, and has a heavy weight attached to its farther end. The boy walks along the side AC, taking with him his end of the rope. Assuming that the instantaneous motion of the weight is always in the direction in which the rope extends away from it, find the differential equation of the path of the weight.

10. A point, P, moves in such a way that it traces a curve the slope of which is proportional to the area of the triangle formed by the ordinate to the point, the x-axis, and the line drawn between the points $(2, 0)$ and P. Find the differential equation of the curve.

11. A machine is accelerated by a constant force F, but is retarded by frictional forces at the bearings, these forces being directly proportional to its velocity, the factor of proportionality being R. However, owing to the heating of the lubricant in the bearings the factor R itself varies with time, and may be taken roughly as the reciprocal of $t + 1$. Remembering that the acceleration of a body is the quotient of the applied force by the mass, set up the differential equation of motion of this machine.

CHAPTER IV

METHODS OF SOLUTION; FIRST ORDER EQUATIONS

§ 21. *Dependent Variable Missing*

Two very simple types of first order equations are those from which one or the other of the variables is missing. Suppose it is the dependent variable y which does not explicitly occur. The equation must then be of the form

$$f\left(\frac{dy}{dx}, x\right) = 0.$$

When this equation is solved for $\frac{dy}{dx}$ an equation such as

$$\frac{dy}{dx} = \phi(x)$$

results. In this form, the problem of solving the differential equation is reduced to a mere matter of integration. The solution is

$$y = \int \phi(x)\, dx + \alpha, \tag{51}$$

α being arbitrary.

It is frequently more convenient to write the solution in the form

$$y = \int_{x_0}^{x} \phi(x)\, dx, \tag{52}$$

in which x_0 is the constant of integration. That this form is identical with (51) is easily seen: for the *definite* integral in (52) may be evaluated by substituting the two limits in *any* (that is, *an indefinite*) integral, and subtracting the results. But when the upper limit is so substituted, it gives only the

indefinite integral itself, while when the lower limit is substituted it gives some constant, the value of which is arbitrary so long as x_0 is arbitrary. If this constant is written $-\alpha$, (52) reduces to (51).

As an example, consider the equation

$$\left(\frac{dy}{dx}\right)^2 + 2\frac{dy}{dx} + x = 0. \tag{53}$$

This equation may be solved for $\frac{dy}{dx}$, with the result

$$\frac{dy}{dx} = -1 \pm \sqrt{1 - x},$$

of which the integral is

$$y = \alpha - x \pm \tfrac{2}{3}(1 - x)^{3/2}.$$

Hence (53) is satisfied when and only when one or the other of the relations

$$\left.\begin{aligned} y + x - \alpha + \tfrac{2}{3}(1 - x)^{3/2} = 0, \\ y + x - \alpha - \tfrac{2}{3}(1 - x)^{3/2} = 0, \end{aligned}\right\} \tag{54}$$

is true.

Both of these solutions can be included in a single equation; for if the product of two quantities vanishes, either one or the other of the quantities must be zero, and conversely. Hence every solution of (53) satisfies the relation

$$(y + x - \alpha)^2 - \tfrac{4}{9}(1 - x)^3 = 0, \tag{55}$$

obtained by multiplying the two equations (54) together term by term. Conversely, whenever (55) is satisfied one or the other of the equations (54) must be satisfied, which means that (53) is satisfied too.

If in addition to the knowledge that y must satisfy (53) it is known that y must vanish at $x = 0$, the value of α is no longer arbitrary, but must be so chosen that this condition will be fulfilled. Upon making both y and x zero in (55) it is found

that $\alpha = \pm \frac{2}{3}$. Hence the particular solution [1] which satisfies both the equation and its boundary condition is

$$(y + x \pm \tfrac{2}{3})^2 = \tfrac{4}{9}(1 - x)^3.$$

§ 22. *Independent Variable Missing*

In case the independent variable is missing the equation takes the form

$$f\left(\frac{dy}{dx}, y\right) = 0.$$

Solved for $\dfrac{dy}{dx}$ this gives

$$\frac{dy}{dx} = \phi(y),$$

or

$$\frac{dy}{\phi(y)} = dx;$$

the solution of which is obviously

$$x = \int \frac{dy}{\phi(y)} + \alpha.$$

It may also be written in the form

$$x = \int_{y_0}^{y} \frac{dy}{\phi(y)}.$$

As an example, consider the equation

$$\sin \frac{dy}{dx} = 1 - y.$$

This reduces to the form

$$\frac{dy}{\sin^{-1}(1 - y)} = dx,$$

or

$$x = \int \frac{dy}{\sin^{-1}(1 - y)} + \alpha. \tag{56}$$

[1] Or, better, the particular solutions; for the two values of α give different curves.

So far as the theory of differential equations is concerned, this is the solution of the problem at hand. The fact that the integral cannot be expressed in finite form in terms of the elementary functions is not gratifying, but it cannot be avoided.[1]

§ 23. *Variables Separable*

The general equation of the first order can contain only the variables x and y and the one derivative $\frac{dy}{dx}$. Theoretically it is always possible, therefore, to solve it for $\frac{dy}{dx}$, and when this is done an equation of the form

$$\frac{dy}{dx} = \phi(x, y) \tag{57}$$

is obtained. If $\phi(x, y)$ happens to be the product of a function of x only and a function of y only, in which case it can be written as $\phi_1(x)\phi_2(y)$, or if it can in some manner be thrown into that form, (57) can be rewritten as

$$\frac{dy}{\phi_2(y)} = \phi_1(x)\, dx.$$

The solution of this is obviously

$$\int \frac{dy}{\phi_2(y)} = \int \phi_1(x)\, dx + \alpha.$$

This process is known as solving a differential equation "by separation of variables." The methods explained in §§ 21 and 22 are special cases of this scheme, and other examples are not difficult to find. For instance, when the variables in (34) are separated, it becomes

$$\frac{dy}{\sqrt{1 - y^2}\sin^{-1} y} = \frac{dx}{x}. \tag{58}$$

[1] In §§ 24 to 26 practical methods of evaluating the integral will be explained. By means of these it is possible, either to obtain a curve representing the relation between x and y, or else to compute a table of values, thereby completing the solution of the problem.

As $\dfrac{dy}{\sqrt{1 - y^2}}$ is the differential of $\sin^{-1}y$, (58) may be written in the alternative form

$$\frac{d \sin^{-1}y}{\sin^{-1}y} = \frac{dx}{x} \tag{59}$$

the solution of which is obviously $\log \sin^{-1}y = \log x + \alpha$. This can readily be reduced to the form (32), from which the differential equation (34) was originally derived.

It is interesting to note that the substitution suggested in Problem 6, § 5, reduced this equation to the form

$$\frac{dw}{dx} = \frac{w}{x}, \tag{60}$$

in which the variables are easily separated. After the integrations have been performed the solution is found to be

$$w = cx;$$

and this in turn reduces to the form

$$y = \sin cx$$

by virtue of the relation between y and w.

Essentially there is no difference between these two methods of solution. The change of variable in Problem 6 was made in accordance with the equation $y = \sin w$, or $w = \sin^{-1}y$. In (59) this same function and its differential were explicitly written, and used as if $\sin^{-1}y$ were itself the variable of integration. Whether or not it is kept explicit as in (58) and (59), or replaced by a w as in (60), is purely a matter of convenience.

PROBLEMS

1. Solve equation (39). Assuming that arc lengths are measured from the minimum point of the curve, determine the value of the constant of integration.

2. Solve Problem 3, § 5. Is the solution the same as the one obtained before?

3. Solve the differential equation derived in Problem 4, § 20. If the temperature of the hot medium is 30° higher than that of the cooler one, and if after two seconds the reading of the thermometer is 20° higher than the temperature of the surrounding medium, what is it twenty seconds after removal? Would this make a good clinical thermometer?

4. Solve the following equations:

(*a*) $\sin x \cos^2 y \, dx + \cos^2 x \, dy = 0.$

(*b*) $\dfrac{dy}{dx} + \sqrt{\dfrac{1 - y^2}{1 - x^2}} = 0.$

(*c*) $y - x\dfrac{dy}{dx} = b\left(1 + x^2 \dfrac{dy}{dx}\right).$

5. Solve equation (36).

6. Solve

$$\frac{dy}{dx} = 1 + \frac{1}{x} - \frac{1}{y^2 + 2} - \frac{1}{x(y^2 + 2)}.$$

§ 24. *Numerical Integration*

The types of equations discussed in §§ 21 to 23 all reduce to the form

$$\int f_1(y) \, dy = \int f_2(x) \, dx + \alpha. \tag{61}$$

Though this constitutes a "classroom solution" in the theory of differential equations, it may fall short of the requirements of the scientist for either of two reasons: he may not be able to evaluate the integrals; or, after they have been evaluated, he may not be able to solve for y as an explicit function of x. Equation (56) is an illustration of the first type of difficulty, while the solution of Problem 6, § 23, will be found to be an excellent example of the other. When either of these conditions arises, some further tools are necessary.

It will serve present purposes best to take these difficulties up in the reverse of their logical order, considering first what is to be done if the functions have been integrated, and then how to integrate them.

If both functions have been integrated — for the moment it does not matter how — the two integrals in (61) can at least be plotted to the same set of axes, as in Fig. 18, the abscissæ being called x for the one curve and y for the other. Then suppose that the boundary condition is so stated that $y = y_0$ when $x = x_0$. These points having been located on the axis, as in Fig. 18, the ordinates corresponding to them represent the values which the integral terms in (61) must take simul-

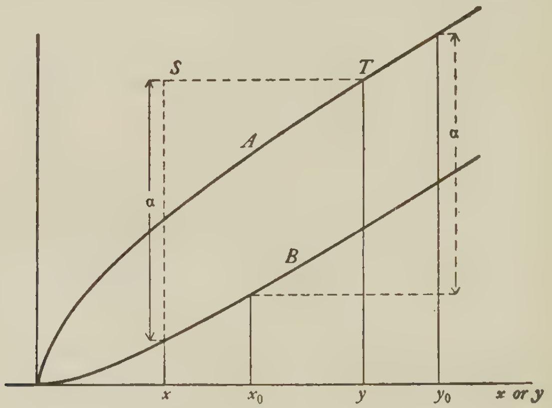

Fig. 18.

taneously. Hence α must have a value equal to the difference in their lengths, for otherwise (61) would not be a true equation.

This value of α being known, it is easy to pick out other pairs of values of x and y; for (61) states that if x and y are any pair of *corresponding* values, the ordinate to the curve A at y must exceed the ordinate to the curve B at x by an amount α. Suppose, then, that we wish to know the value of y which corresponds to any x. This x having been marked on the axis of abscissæ, and the amount α having been added to the ordinate of curve B, thus extending it to the point S, we need only seek the point T on curve A which is at the same height as S, and the scale-reading immediately below it will be the y for which we seek.

Other values of x may be chosen, and the corresponding values of y determined by the same process. When a suf-

ficient number are available, they may be plotted as in Fig. 19 and a curve C drawn through them. This curve then represents the relationship between x and y. It may obviously be extended as far as we desire.

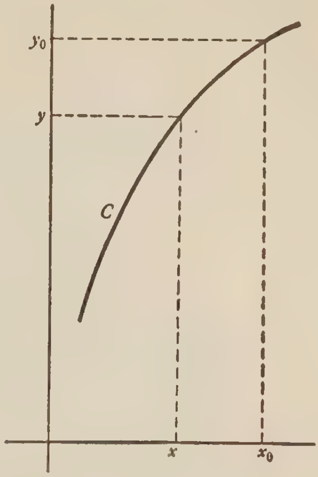

FIG. 19.

§ 25. *Integration in Series*

The method of § 24 presupposes the possibility of integrating $f_1(y)$ and $f_2(x)$. The next few sections will be devoted to explaining methods which are sometimes useful for this purpose.

The functions $f_1(y)$ and $f_2(x)$ can often be expanded in Taylor's series. If the series converge in the neighborhood of the values y_0 and x_0, they can be integrated term by term and computations may be made from the series which result.

As an example, consider the equation

$$x = \int \frac{dy}{\sin^{-1}(1 - y)} + \alpha, \tag{56}$$

which appeared in § 22. To simplify matters somewhat, y will be replaced by a new variable w defined by the equation

$$w = \sin^{-1}(1 - y). \tag{62}$$

Then

$$x = -\int \frac{\cos w}{w}\, dw + \alpha.$$

This, of course, can be written in the alternative form

$$x = -\int_{w_0}^{w} \frac{\cos w}{w}\, dw. \tag{63}$$

Now it is not possible to integrate $\dfrac{\cos w}{w}$ in terms of the elementary functions; if it were, (56) could also be integrated

at once. But it is possible to expand it in a rapidly converging series, for the series

$$\cos w = 1 - \frac{w^2}{2!} + \frac{w^4}{4!} - \frac{w^6}{6!} + \cdots$$

is well known, and from it we obtain at once

$$\frac{\cos w}{w} = \frac{1}{w} - \frac{w}{2!} + \frac{w^3}{4!} - \frac{w^5}{6!} + \cdots.$$

When this is integrated term by term it gives

$$x = \alpha - \log w + \frac{w^2}{2 \cdot 2!} - \frac{w^4}{4 \cdot 4!} + \frac{w^6}{6 \cdot 6!} - \cdots. \quad (64)$$

Suppose, now, that the boundary condition is $y = 0$ when $x = 0$. Then by (62) $w = \sin^{-1} 1 = \pi/2$. Substituting this value in the right-hand side of (64) gives the value $\alpha + 0.1052$. Hence, if x is to be zero when y is zero, α must have the value -0.1052. The relation between x and y is now completely determined by means of the equations (62) and (64). Since (64) is convergent for *any* value of w, it is possible to compute x for any value of y and plot a curve relating them to one another. From the standpoint of computation, however, it would be simpler to compute x for suitable values of w^2 by the use of (64), and to find the corresponding values of y from (62); for in doing so, the w's might be so chosen as to have only a small number of digits, so that their logarithms could be found without interpolation, and at the same time the powers needed in (64) would be more easily computed.

The integral

$$\int_{w_0}^{w} \frac{\cos w}{w}\, dw,$$

which occurs in (63) is frequently met in studies on the diffraction of light and sound. Its values have been extensively tabulated under the name "Integral Cosine," and are usually represented by the symbol[1] Ci w. Of course, the tables

[1] The symbol is analogous to "cos w" and the like, and is derived from the initial letters of the Latin words for "integral cosine." The inverted order is due to the fact that in Latin the adjective usually follows the noun.

represent a *definite*, not an *indefinite* integral. That is, they correspond to a particular value of w_0, which happens to be ∞. Hence in expressing (63) in terms of these tabulated functions we would have to write

$$x = \mathrm{Ci}\, w_0 - \mathrm{Ci}\, w. \tag{65}$$

It is not hard to determine w_0 so as to satisfy our boundary value, for since x must vanish when $w = \pi/2$, it is obvious that w_0 must be $\pi/2$. Hence

$$x = \mathrm{Ci}\, \frac{\pi}{2} - \mathrm{Ci}\, w$$

is a shorthand form of (64) and corresponds to the particular value $\alpha = -0.1052$.

The fact that the simple expression (65) and the series (64) represent one and the same function may serve to illustrate how slender is the distinction between those functions which are usually thought to be simple and those which are regarded as involved. When first met, Ci w is likely to appear as an artificial notation for the complicated series (64) — not as being at all in the same class with such "simple" functions as $\log w$, $\cos w$, and e^w. Yet these also are available for purposes of computation only because extensive tables exist, and those tables were originally derived from series.

Similarly, there is no essential reason for saying, as might be done in the study of the Calculus, that the integral

$$\int \frac{dw}{w}$$

can be evaluated, while the integral

$$\int \frac{\cos w}{w}\, dw$$

cannot: for one integral may be written $\log w$ and the other Ci w, each of which is available as a tabulated function and in no other way.

There is, therefore, nothing esoteric in the use of series solutions. If desired, the series solution of any integral could

be given a name, as (63) has been given the name Ci w, and its values could be tabulated. It would then be placed in just the same category as logarithms and trigonometric functions. That this has been done in certain cases and not in others is due to different degrees of usefulness, not to any other characteristic.

Taylor's series, moreover, is not necessarily the best form in which to expand a function. Thus, while (64) converges rapidly for small values of w, it is hard to deal with when w is large, and other series can easily be found which serve the purpose better. Such a series will next be derived; not for the general integral, however, but for Ci w itself, since we have seen that the general integral is easily expressed in terms of it by means of (65).

Upon integrating

$$\operatorname{Ci} w = \int_{\infty}^{w} \frac{\cos w}{w}\, dw$$

by parts, we obtain

$$\operatorname{Ci} w = \frac{\sin w}{w} + \int_{\infty}^{w} \frac{\sin w}{w^2}\, dw.$$

Integrating this result by parts a second time, we get

$$\operatorname{Ci} w = \frac{\sin w}{w} - \frac{\cos w}{w^2} - 2\int_{\infty}^{w} \frac{\cos w}{w^3}\, dw.$$

Continuing this process, we are led to the series

$$\operatorname{Ci} w = \sin w \left[\frac{1}{w} - \frac{1\cdot 2}{w^3} + \frac{1\cdot 2\cdot 3\cdot 4}{w^5} - \frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6}{w^7} + \cdots\right]$$
$$- \cos w \left[\frac{1}{w^2} - \frac{1\cdot 2\cdot 3}{w^4} + \frac{1\cdot 2\cdot 3\cdot 4\cdot 5}{w^6} - \cdots\right]. \qquad (66)$$

This series does not converge, but nevertheless it is useful for purposes of computation when w is large; for it may be shown that if n terms of the series are computed, their sum differs from the true answer by an amount less than $\dfrac{1\cdot 2\cdot 3 \cdots (n-1)}{w^n}$. To see this, we note that after n integra-

tions by parts, Ci w will be expressed exactly as the sum of the first n terms of (66), plus an integral which is either

$$+\ 1\cdot 2\cdot 3\cdots n\int_{\infty}^{w}\frac{\sin w}{w^{n+1}}\,dw$$

or

$$-\ 1\cdot 2\cdot 3\cdots n\int_{\infty}^{w}\frac{\cos w}{w^{n+1}}\,dw.$$

One or the other of these terms, therefore, represents the error incurred upon accepting the first n terms of (66) as a sufficient approximation. But since neither sin w nor cos w ever exceeds 1, these errors must necessarily be smaller in absolute value than

$$1\cdot 2\cdot 3\cdots n\int_{w}^{\infty}\frac{1}{w^{n+1}}\,dw=\frac{1\cdot 2\cdot 3\cdots(n-1)}{w^{n}}.$$

If, then, the terms inside the parentheses in (66) first *decrease* in value and become very small — as they actually do when w is large — and afterward *increase* and cause the series to diverge, a good approximation may be obtained by stopping the computation while the terms are still small. To obtain the same degree of accuracy from (64) would require much more labor.

PROBLEMS

1. Find a series expansion for (176), § 59, by expanding e^{-x} in a power series.

2. Find, by repeated integration by parts, an expansion for the same function in *descending* powers of x.

3. Compute the values of Ci 1 and of Ci 10 to four decimal places, given that Ci $\pi/2 = 0.4720$.

§ 26. *Graphical Integration*

When the uses to which the solution of an equation are to be put are such that no very great accuracy is demanded, a process of graphical integration can sometimes be used to considerable advantage. The method which is most feasible makes use of two ideas. The first is that the integral of a constant is represented graphically by a straight line the slope

of which is equal to that constant. The second is that the integral $\int_a^b f(x)\,dx$ is equal to the area bounded by the curve $y = f(x)$, the x-axis, and the ordinates at $x = a$ and $x = b$.

Suppose it is required to integrate the function $y = f(x)$, which is represented in Fig. 20 by the curve $ABC \cdots$; and suppose that this curve is replaced by the broken line $abcd \cdots$, so constructed that the two "triangles" of each of the pairs α, α; β, β; $\cdots$; are equal in area. The integral of the

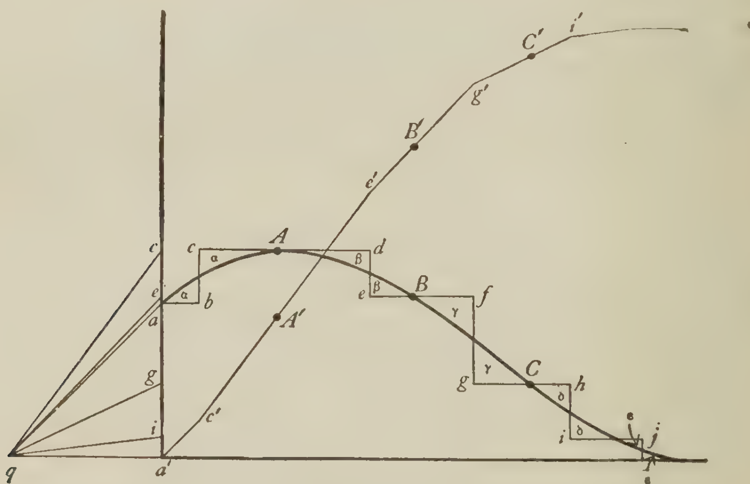

FIG. 20. — GRAPHICAL INTEGRATION.

broken line can easily be obtained by the use of the first idea mentioned above: that is, by drawing a succession of straight lines with appropriate slopes. Suppose this has been done, and has led to the broken line $a'c'e' \cdots$. At the points $A', B', \cdots$, of this broken line the ordinates must be equal to the integral of the function $f(x)$ itself; for up to these ordinates the areas under $ABC \cdots$ and its approximation $abcd \cdots$ are equal. The true integral curve therefore passes through the points $A', B', C', \cdots$. Furthermore, at the points $A, B, C, \cdots$, to which $A', B', C', \cdots$, correspond, both the line segments $cd, ef, \cdots$, and the true curve ABC have the same ordinates. Hence at these places the true

integral curve should have the same slope as the broken line $a'c'e' \cdots$. Thus not only does the true curve pass through $A', B', \cdots$, but it is tangent to the broken line at these points. The knowledge of these facts makes it possible to draw in a curve which represents $\int f(x)\,dx$ to a high degree of approximation.

A practical routine for carrying out this process is illustrated in Fig. 20.

The broken line is first constructed in such a way that, as nearly as can be estimated by eye, the corresponding triangles

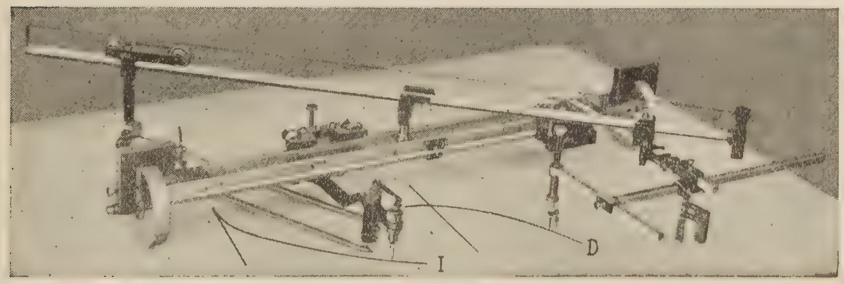

FIG. 21. — THE INTEGRAPH.

of each pair are equal. Points $a, c, e, \cdots$, are then marked on the y-axis at heights corresponding to the segments $ab, cd, ef, \cdots$, respectively, and a point q is marked on the x-axis, one unit to the left of the origin. The lines $qa, qc, qe, \cdots$, therefore have the same slopes as the segments $a'c', c'e', e'g', \cdots$. The curve $A'B'C' \cdots$ was, in fact, constructed by joining together segments parallel to these lines.

§ 27. *The Integraph*

There is a machine, manufactured by a Swiss instrument-maker,[1] which is so constructed that, when a pointer is passed along the curve $y = f(x)$, a pen associated with the machine automatically draws the curve $y = \int f(x)\,dx$. It is shown in

[1] G. Coradi, Zurich.

Fig. 21, and in Fig. 22 is given a schematic diagram illustrating its method of operation.

The essential feature of the machine is a triangle ABC, having three movable sides. The side AB is a heavy carriage, which is kept parallel to the y-axis, but which rolls freely on the wheels WW. The side BC is kept parallel to the x-axis, but is supported in such a way that it may move freely in the y-direction. Its length, which is maintained fixed, is the theoretical unit of length for the operation of the machine. The third side AC is a rod so supported as to pass through the point A at which the x-axis intersects the carriage, and through the end C of the rod BC. At C is placed a pointer,[1] by means of which to trace out the curve $y = f(x)$ of which the integral is desired.

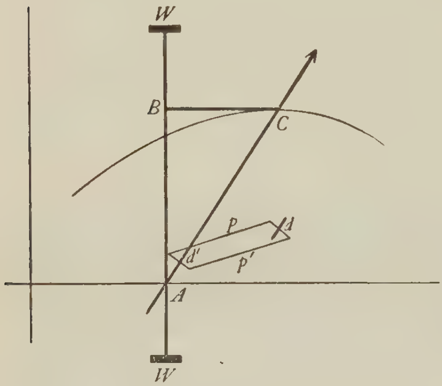

Fig. 22. — Schematic Diagram of the Integraph.

Obviously, since BC is unity, the slope of the line AC is always equal to the value of $f(x)$ which corresponds to the particular point C upon which the tracing point is resting.

There is also a parallelogram $pp'dd'$, the side d' of which is maintained at right angles to the bar AC, though it is free to slide along it. The center of the opposite side d is located on the ordinate to the point C, and is the axis of a knife-edged disc which rests on the paper and rolls as the carriage AB is moved. As its axis d is *perpendicular* to AC its motion must be *parallel* to AC. Hence as the pointer is moved the disc travels a path the slope of which is constantly equal to the instantaneous ordinate $f(x)$ of the point C. Hence if C moves over

[1] For convenience in operation, the pointer is carried by an adjustable framework, as shown in Fig. 21; but this has no bearing on the theory of the machine.

the curve $y = f(x)$, the disc must roll along a path the equation of which is $y = \int f(x)\, dx$.

In the integraph a pen is associated with the disc so as to draw this integral curve, thus leaving a permanent record of the integration.

Institutions where many differential equations are met, so that difficult integrations are frequent, usually possess such devices. Their accuracy is quite sufficient for most practical purposes, and they are rapid in operation. They possess one disadvantage, however, which is frequently serious. If the differential equation contains literal constants or parameters, instead of numerical ones, it may be necessary to assign a succession of values to these constants and carry out a separate solution for each. This, of course, requires a great deal of labor, and may prove prohibitive.

For example, the equation

$$\frac{dy}{dx} = x \sin (3y^2 + 2)$$

could easily be solved by means of the integraph; but the equation

$$\frac{dy}{dx} = x \sin (ay^2 + b) \tag{67}$$

could only be solved by assigning special values to the constants a and b.

On the other hand such parameters may frequently be caused to disappear by an appropriate change of variable. Thus, the equation

$$\frac{dy}{dx} = x\,(ay^2 + b) \tag{68}$$

appears at first sight to be as difficult as the other. But by introducing two new variables

$$x = c_1\xi,$$

$$y = c_2\eta,$$

it may be caused to take the form

$$\frac{d\eta}{d\xi} = \frac{c_1^2}{c_2}(ac_2^2\eta^2 + b)\xi.$$

Then, by making c_1 and c_2 satisfy the conditions $ac_1^2 c_2 = 1$ and $\frac{bc_1^2}{c_2} = 1$, $\left(\text{that is, by making } c_1 = \frac{1}{\sqrt[4]{ab}} \text{ and } c_2 = \sqrt{\frac{b}{a}}\right)$, this may be reduced to the simpler form

$$\frac{d\eta}{d\xi} = (\eta^2 + 1)\xi. \tag{69}$$

Hence by solving (69) a curve connecting ξ and η could be obtained, such as that shown in Fig. 23. As this curve applies equally well to *all* values of a and b, the solution of *any* equation of the form (68) can be obtained from it by merely changing the scales on the axes. Thus, if $b = 8$ and $a = 2$, x and y are related to ξ and η by the laws $x = \frac{1}{2}\xi$ and $y = 2\eta$. Therefore, by multiplying the numbers on the η-axis by 2 and dividing those on the ξ-axis by 2 the $\xi\eta$-curve is transformed into one connecting x and y.

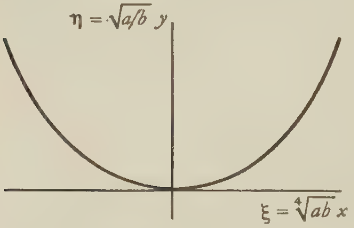

FIG. 23.

Even (67) is capable of considerable simplification by this device, for by using the substitutions $\eta = y\sqrt{a}$ and $\xi = x\sqrt[4]{a}$ it may be reduced to the form

$$\frac{d\eta}{d\xi} = \xi \sin(\eta^2 + b),$$

which contains only one parameter instead of two.

§ 28. *Variables Separable by Substitution; Homogeneous Equations*

It sometimes happens that an equation in which the variables do not appear to be separable may be reduced to a form in which they are separable by the use of a suitable substitution.

As an example, consider the equation

$$x\frac{dy}{dx} + y + y^2x = 0.$$

Upon replacing y by $w = xy$ we obtain a new equation

$$\frac{dw}{dx} + \frac{w^2}{x} = 0,$$

of which the solution is

$$x = \alpha e^{\frac{1}{w}} = \alpha e^{\frac{1}{xy}}.$$

The reader will no doubt feel that this is a piece of trickery which is only likely to work in problems that have been hand-picked for the purpose. There is, indeed, a certain amount of truth in this reaction, for even where a suitable substitution exists it is not likely to be obvious.[1] Consequently this method of solution should generally be one of the last to be tried. There is one type of equation, however, for which the correct substitution is always known. It is known as the *homogeneous equation*.

An equation

$$\frac{dy}{dx} = \phi(x, y) \tag{70}$$

is said to be homogeneous if the function ϕ remains unaltered when x and y are replaced by kx and ky, regardless of the value of the quantity k. That is, the equation is homogeneous if

$$\phi(x, y) \equiv \phi(kx, ky)$$

for any value of k whatsoever. In particular, if k is replaced by

[1] In the illustration, the substitution might be suggested by the observation that the first two terms are just the well-known formula for $\frac{d}{dx}(xy)$. It often happens, in dealing with troublesome equations, that success or failure hangs upon just such an observation as this. The solution of differential equations is no fit occupation for the obtuse.

$1/x$, the arguments of ϕ become 1 and y/x. Then (70) takes the form

$$\frac{dy}{dx} = \phi\left(1, \frac{y}{x}\right).$$

This suggests substituting $w = y/x$ for y, which is found to lead to an equation

$$x\frac{dw}{dx} + w = \phi(1, w)$$

in which the variables are separable.

From this point on the formal process is:

$$\frac{dw}{\phi(1, w) - w} = \frac{dx}{x};$$

$$\log x = \log \alpha + \int \frac{dw}{\phi(1, w) - w};$$

$$x = \alpha e^{\int \frac{dw}{\phi(1, w) - w}}.$$

As an example, consider the equation

$$(x^2 + y^2)\frac{dy}{dx} + 2xy = 0, \tag{71}$$

or

$$\frac{dy}{dx} = -\frac{2xy}{x^2 + y^2}.$$

It is easily seen that, if kx and ky are substituted for x and y, no change occurs in the right-hand side of the equation. Hence it is homogeneous. Now set $y = wx$. Then

$$x\frac{dw}{dx} + w = -\frac{2w}{1 + w^2},$$

or

$$\frac{1 + w^2}{3w + w^3}\,dw = -\frac{dx}{x}.$$

The integral of this is

$$-\log x = \log \alpha + \tfrac{1}{3}\log(3w + w^3),$$

which reduces to

$$x^2y + \frac{y^3}{3} = \frac{1}{3\alpha^3}.$$

It may be remarked in passing that this reduction to a form in which the variables are separable can be brought about by replacing x by w equally as well as by using the substitution made above. However, it is usually simpler to change the dependent variable than to change the independent one, and consequently the substitution of w for y is the one most commonly used.

§ 29. *Exact Differential Equations*

Some first order differential equations may be solved by comparison with the formula for total differentiation

$$\frac{d\phi}{dx} = \frac{\partial\phi}{\partial x} + \frac{\partial\phi}{\partial y}\frac{dy}{dx}. \tag{72}$$

Suppose that the differential equation to be solved occurs in, or can be reduced to, the form

$$f_1(x, y) + f_2(x, y)\frac{dy}{dx} = 0. \tag{73}$$

Obviously, if we could find a function $\phi(x, y)$ such that $\dfrac{\partial\phi}{\partial x} = f_1$ and $\dfrac{\partial\phi}{\partial y} = f_2$, (72) and (73) would become identical upon the assumption that $\dfrac{d\phi}{dx} = 0$. It would therefore follow at once that the solution of (73) was $\phi(x, y) = \alpha$, since otherwise $\dfrac{d\phi}{dx}$ would not vanish.

This, of course, does not really constitute a method for solving such equations; for it does not tell us how to find the function ϕ. It may, however, be made the basis of a method of solution, provided two things can be done: first, provided

we can invent a method for determining from the equation itself whether or not such a function ϕ exists, and second, provided we can invent a method for finding the function if it does exist. Actually, both these things can be done, as we shall now see. First, however, we must remark that the equation (73) is called *exact* when such a function ϕ exists, so that our first object may be restated as *a criterion for determining whether an equation is exact.*

To find such a criterion, suppose (73) were exact. Then, by definition,

$$\frac{\partial \phi}{\partial x} = f_1,$$

$$\frac{\partial \phi}{\partial y} = f_2.$$

Differentiating the first of these with respect to y, and the second with respect to x, it is seen that

$$\frac{\partial^2 \phi}{\partial y \partial x} = \frac{\partial f_1}{\partial y},$$

$$\frac{\partial^2 \phi}{\partial x \partial y} = \frac{\partial f_2}{\partial x};$$

from which it follows [1] that, *if the equation is exact*

$$\frac{\partial f_1}{\partial y} = \frac{\partial f_2}{\partial x}. \tag{74}$$

Moreover, the converse of this theorem can also be proved: namely, that if (74) is satisfied (73) is exact. Thus it is always possible to decide whether or not an equation is exact *before attempting to solve it.*

[1] There are exceptional functions for which $\frac{\partial^2 \phi}{\partial x \partial y}$ is not equal to $\frac{\partial^2 \phi}{\partial y \partial x}$; but these are of such rare occurrence, particularly in the applied sciences, that they can quite properly be ignored in an elementary textbook.

To illustrate: consider again the example

$$(x^2 + y^2)\frac{dy}{dx} + 2xy = 0. \tag{71}$$

Here $f_1 = 2xy$ and $f_2 = x^2 + y^2$. Hence

$$\frac{\partial f_1}{\partial y} = \frac{\partial f_2}{\partial x} = 2x,$$

and the equation is exact.

Having now accomplished our first objective, we turn to the matter of finding the function ϕ, which we shall find to be equally simple. Once an equation is known to be exact the relations

$$\frac{\partial \phi}{\partial x} = f_1(x, y)$$

and

$$\frac{\partial \phi}{\partial y} = f_2(x, y)$$

are known to be true. Since $\dfrac{\partial \phi}{\partial x}$ is gotten from ϕ by ordinary differentiation, *assuming y to be constant*, it follows that ϕ may be obtained from $\dfrac{\partial \phi}{\partial x}$ by ordinary integration, *assuming y to be constant.* Hence

$$\phi = \int f_1(x, y)\,\partial x + \alpha(y), \tag{75}$$

the element of integration being written ∂x instead of dx to indicate that y is to be considered constant, and the "constant of integration" being written $\alpha(y)$ since, so far as partial differentiation with respect to x is concerned, any function of y only is to be so regarded.

Integrating the second equation in the same way, we get

$$\phi = \int f_2(x, y)\,\partial y + \beta(x). \tag{76}$$

Upon comparing (75) and (76) two possibilities present

themselves: In the first place, $\int f_2(x, y)\,\partial y$ may be equal to $\int f_1(x, y)\partial x$, in which case the two equations will define the same function ϕ if and only if $\alpha(y)$ and $\beta(x)$ are equal, which of course requires them both to be constant.

In the second place, $\int f_1(x, y)\,\partial x$ may contain a function of x, say $g(x)$, which does not appear in $\int f_2(x, y)\,\partial y$; or $\int f_2(x, y)\,\partial y$ may contain a function of y, say $h(y)$, which does not appear in $\int f_1(x, y)\,\partial x$, or both. In this case, in order that (75) and (76) shall be identical it is necessary that $\beta(x) = g(x)$ and $\alpha(y) = h(y)$. In either event the solution of the equation is known to be $\phi = C$, C being a constant.

As an example, consider the equation (71) dealt with above. Since

$$\frac{\partial\phi}{\partial x} = 2xy,$$

it follows that

$$\phi = \int 2xy\,\partial x + \alpha(y)$$

$$= x^2y + \alpha(y).$$

Also, since

$$\frac{\partial\phi}{\partial y} = x^2 + y^2,$$

$$\phi = x^2y + \frac{y^3}{3} + \beta(x).$$

Comparing these, they are seen to be identical provided $\beta(x) = 0$ and $\alpha(y) = y^3/3$. Hence $\phi = x^2y + y^3/3$ and the solution of the equation is

$$x^2y + \frac{y^3}{3} = C.$$

The same solution was obtained by another method in § 28.

As a second example, consider the equation

$$\sec^2 x \tan y\, dx + \sec^2 y \tan x\, dy = 0.$$

Here

$$f_1(x, y) = \sec^2 x \tan y,$$

$$f_2(x, y) = \sec^2 y \tan x.$$

Hence

$$\frac{\partial f_1}{\partial y} = \frac{\partial f_2}{\partial x} = \sec^2 x \sec^2 y.$$

Integrating as explained above it is found that

$$\phi = \tan x \tan y + \alpha(y)$$

$$\phi = \tan x \tan y + \beta(x).$$

Hence $\alpha = \beta = C$, a constant, and the solution is [1]

$$\tan x \tan y + C = 0.$$

PROBLEMS

1. $y^2 = x(y - x)\dfrac{dy}{dx}$.

2. $2x^2y + y^3 - x^3\dfrac{dy}{dx} = 0$.

3. $(2ax + by) + (2cy + bx + e)\dfrac{dy}{dx} = g$.

4. $\sec^2 x \tan y \dfrac{dy}{dx} + \sec^2 y \tan x = 0$.

5. $t + I\dfrac{dI}{dt} = mI$.

6. $2\dfrac{\theta}{r^3} + \left(\dfrac{1}{r^2} - 3\dfrac{\theta^2}{r^4}\right)\dfrac{dr}{d\theta} = 0$.

7. $(a\, d\alpha + b\, d\beta) \sin (a\alpha + b\beta) = (b\, d\alpha + a\, d\beta) \cos (b\alpha + a\beta)$.

8. $\left(T + \dfrac{1}{\sqrt{t^2 - T^2}}\right)\dfrac{dT}{dt} = \left(\dfrac{T}{t\sqrt{t^2 - T^2}} - t\right)$.

---

[1] The rule is that ϕ must be set equal to a constant. But there is already one arbitrary constant C, and another would add nothing in the way of generality.

§ 30. *Linear Equations*

If an equation of the first order is linear, it can be thrown into the form

$$\frac{dy}{dx} + f_1(x)\, y = f_2(x). \tag{77}$$

Problem 3, § 5, was of this type, and it was found that when $w = e^{cx}y$ had been substituted for y, it was easily solved. This suggests that a solution might be found in the present instance also by substituting in place of y some new variable $w = y\, g(x)$, provided we were to choose the function $g(x)$ properly. As the proper choice is not at all obvious, however, we shall proceed with the function undetermined, in the hope that something may occur to suggest a suitable choice.

Our substitution $w = gy$ transforms (77) into

$$\frac{dw}{dx} + \left(f_1 - \frac{1}{g}\frac{dg}{dx}\right) w = g f_2.$$

If, now, we choose g so that

$$\frac{1}{g}\frac{dg}{dx} = f_1$$

the second term of this equation disappears, and we have at once

$$w = \alpha + \int g f_2\, dx.$$

But obviously the condition which we have placed upon g leads to the choice

$$g = e^{\int f_1 dx},$$

whence we have

$$w = \alpha + \int e^{\int f_1 dx} f_2\, dx.$$

Changing from w to y, the solution of (77) is found to be

$$y = e^{-\int f_1 dx}\left(\alpha + \int e^{\int f_1 dx} f_2\, dx\right). \tag{78}$$

This formula (78) should be remembered, since many of the equations which can be solved by it can be solved in no other way. In order that the various steps in its derivation may be perfectly clear, the solution of the equation

$$\frac{dy}{dx} + xy = x, \tag{79}$$

will be carried through step by step. Here

$$f_1(x) = x,$$
$$f_2(x) = x.$$

Hence the substitution to be made is

$$w = ye^{\int x\,dx} = ye^{\frac{x^2}{2}}.$$

From this it is found that

$$\frac{dy}{dx} = e^{-\frac{x^2}{2}}\frac{dw}{dx} - xy;$$

whence the original equation becomes

$$\frac{dw}{dx} = xe^{\frac{x^2}{2}}.$$

The solution is therefore

$$w = e^{\frac{x^2}{2}} + \alpha,$$

or

$$y = 1 + \alpha e^{-\frac{x^2}{2}}.$$

As another example, consider the equation

$$\frac{dy}{dx} + \frac{y}{x} = \sin x. \tag{80}$$

Here $f_1 = \frac{1}{x}$ and $f_2 = \sin x$. Hence $\int f_1\,dx = \log x$ and (78) becomes

$$y = e^{-\log x}\left(\alpha + \int e^{\log x}\sin x\,dx\right).$$

But by the definition of a logarithm, $e^{\log x} = x$ and $e^{-\log x} = \frac{1}{x}$. Therefore

$$y = \frac{1}{x}\left(\alpha + \int x \sin x \, dx\right),$$

which is easily evaluated.

§ 31. *Equations Reducible to the Linear Type*

Some equations which are not linear can be made so by an appropriate change of variable. Perhaps the most important type is

$$\frac{dy}{dx} + f_1(x)y = f_2(x)y^m.$$

If this equation is divided by $\dfrac{y^m}{1 - m}$ it takes the form

$$\frac{1 - m}{y^m}\frac{dy}{dx} + (1 - m)f_1 \frac{1}{y^{m-1}} = (1 - m)f_2. \qquad (81)$$

A moment's inspection shows that the first term is the x-derivative of the quantity $\dfrac{1}{y^{m-1}}$, which also occurs in the second term. Hence if this is called w, (81) takes the form

$$\frac{dw}{dx} + (1 - m)f_1 w = (1 - m)f_2,$$

which is linear and easily solved by formula (78). When w is replaced by y^{1-m} in the result it is found that

$$y^{1-m} = e^{-\int (1-m)f_1 dx}\left(\alpha + \int e^{\int (1-m)f_1 dx}(1 - m)f_2 \, dx\right). \qquad (82)$$

Formula (82) differs from (78) only in that each f is replaced by $(1 - m)f$, and y is replaced by y^{1-m}.

As an example, consider the equation

$$\frac{dy}{dx} + \frac{y}{x} = \frac{\sin x}{y^3}. \qquad (83)$$

Here $m = -3$, and therefore the correct substitution is $w = y^4$. Replacing y by this variable it is found that

$$\frac{dw}{dx} + \frac{4}{x} w = 4 \sin x,$$

a linear equation of the type (77) in which $f_1 = \frac{4}{x}$ and $f_2 = 4 \sin x$. Hence the solution is

$$w = y^4 = \frac{1}{x^4}\left(\alpha + 4\int x^4 \sin x \, dx\right),$$

which can readily be evaluated.

PROBLEMS

1. Solve (79) by separation of variables.
2. Evaluate the solution of (80).
3. Evaluate the solution of (83).
4. $\dfrac{d\rho}{d\theta} = \dfrac{\rho + a\theta^3 - 2\rho\theta^2}{\theta(1 - \theta^2)}$.
5. $(T \log t - 1) T \, dt = t \, dT$.
6. $\dfrac{dy}{dx} + y \cos x = \frac{1}{2} \sin 2x$.
7. $y - \cos x \dfrac{dy}{dx} = y^2 \cos x (1 - \sin x)$.
8. $\dfrac{dy}{dx} + y \dfrac{d\phi}{dx} = \phi \dfrac{d\phi}{dx}$.
9. Solve the differential equation set up in Problem 4, § 20, by the method of § 30.

§ 32. *Equations Solvable for x or for y*

There is one more method of attacking first order equations which is worth trying if none of those explained above is successful. That is to solve for one of the variables x or y.

Suppose the equation is solved for x. It is then in the form

$x = \phi(y, y')$, where $y' = \dfrac{dy}{dx}$. Differentiating this, an equation

$$1 = \frac{\partial \phi}{\partial y} y' + \frac{\partial \phi}{\partial y'} \frac{dy'}{dx}$$

is obtained, which contains only the variables y and y', and the derivative $\dfrac{dy'}{dx}$ which may be written $y' \dfrac{dy'}{dy}$. That is, the process above has virtually replaced x by a new variable y'. It *may* be possible to solve this equation and obtain a relation between y' and y. If so, the algebraic elimination of y' between this and the original equation gives the desired solution. If not, the method has failed.

As an example, consider the equation (53), which has already been solved by other means. This becomes

$$x = -y'^2 - 2y', \tag{84}$$

when solved for the variable x. Differentiating and changing $\dfrac{dy'}{dx}$ into $y' \dfrac{dy'}{dy}$, the new differential equation

$$-dy = (2y'^2 + 2y')\,dy'$$

is obtained, the solution of which is obviously

$$-y = \tfrac{2}{3} y'^3 + y'^2 + \alpha. \tag{85}$$

By solving (84) for y', it is found that

$$y' = -1 \pm \sqrt{1 - x},$$

and upon substituting this in (85) the final solution

$$y + \alpha + (\tfrac{1}{3} \pm \tfrac{2}{3}\sqrt{1-x})(2 - x \mp 2\sqrt{1-x}) = 0 \tag{86}$$

is obtained.

In spite of its complicated appearance this equation is really identical with the solution obtained in § 21, and can be reduced to the same form as (55), except for a difference in the values of α in the two equations, which is immaterial.

A similar method of attack rests upon solving the original

equation for y, thus reducing it to the form $y = \phi(y', x)$. Differentiating this with respect to x leads to

$$y' = \frac{\partial \phi}{\partial y'} \frac{dy'}{dx} + \frac{\partial \phi}{\partial x},$$

which contains only y', x and $\frac{dy'}{dx}$, and is therefore a first order differential equation in y'. It *may* be possible to solve this for a relation between y' and x, and if so y' may be eliminated from the original equation by means of this relation.

As an example, consider the equation

$$\frac{dy}{dx} + xy = x, \tag{87}$$

which, when solved for y gives

$$y = 1 - \frac{y'}{x}.$$

Differentiating,

$$y' = -\frac{1}{x}\frac{dy'}{dx} + \frac{y'}{x^2},$$

or

$$\frac{dy'}{y'} = \frac{1 - x^2}{x}\, dx.$$

The solution of this equation is

$$y' = \alpha x\, e^{-x^2/2}.$$

Eliminating y' between this and the original equation the solution is found to be

$$y = 1 - \alpha\, e^{-x^2/2}.$$

The same solution was gotten in § 30, except for the sign of α, which is immaterial since α is arbitrary.

Both of the examples used in this section are easily solved by other means. The following examples, on the contrary, though not difficult when dealt with by the present method, would be very troublesome without it.

The equation

$$x^2y'^2 + y^2 = y'(1 + 2xy), \tag{88}$$

may easily be solved for y, giving

$$y = xy' \pm \sqrt{y'}.$$

Differentiation gives

$$y' = x\frac{dy'}{dx} + y' \pm \frac{1}{2\sqrt{y'}}\frac{dy'}{dx}$$

or [1]

$$\frac{dy'}{dx} = 0.$$

The solution of this equation is $y' = \alpha$, by means of which the general solution of (88) is obtained in the form

$$\alpha^2x^2 + y^2 = \alpha + 2\alpha xy.$$

As a second example, consider the equation

$$\frac{dy}{dx} = (x + y)^2. \tag{89}$$

If this is solved for x, there results

$$x = -y + \sqrt{y'}.$$

Differentiation gives

$$1 = -y' + \frac{1}{2\sqrt{y'}}\frac{dy'}{dx},$$

or

$$dx = \frac{dy'}{2(1 + y')\sqrt{y'}}.$$

---

[1] In searching for the general solution of a differential equation, factors which do not involve derivatives can be neglected in the same way that constants are neglected in solving algebraic equations. On this basis, the factor $x \pm \frac{1}{2\sqrt{y'}}$, by which $\frac{dy'}{dx}$ is multiplied, may be discarded.

The appearance of such algebraic factors is, however, an indication of the existence of "singular solutions"; that is, solutions of the differential equation which are not special cases of the general solution. These will be discussed in Chapter V.

The solution of this is easily found to be

$$y' = \tan^2 (x + \alpha).$$

Hence the general solution of (89) is

$$(x + y)^2 = \tan^2 (x + \alpha).$$

PROBLEMS

1. Reduce (86) to the form (55).

2. $x\left(\frac{dy}{dx}\right)^2 - y + 2\frac{dy}{dx} = 0.$

3. $2\left(\frac{d\phi}{dr}\right)^3 + \left(\frac{d\phi}{dr}\right)^2 - \phi = 0.$

4. $\frac{dy}{dz} = e^{z - \frac{dy}{dz}}.$

5. $\sqrt{t^2 + T} = \frac{dT}{dt}.$

6. $y = x\frac{dy}{dx} + \phi\left(\frac{dy}{dx}\right)$, where ϕ is any function whatsoever.

7. $(x^2 - 1)\left(\frac{dy}{dx}\right)^2 = 1.$

8. Solve (89) by substituting $w = x + y$ in place of y.

§ 33. *Second Order Equations Reducible to First Order Equations*

Many equations which appear to be of higher order than the first become first order equations when y' is written in place of $\frac{dy}{dx}$. For instance, the equation derived in Problem 5, § 20, is of the form $F\left(\frac{dy}{dx}, \frac{d^2y}{dx^2}\right) = 0$. In terms of y' this becomes $F\left(y', \frac{dy'}{dx}\right) = 0$, which is a first order equation the solution of which may easily be found. The result is, of course, a new equation connecting y', y and x; and this itself is a first order differential equation.

There are three types of second order differential equations to which this method is applicable. They are, (*a*) those which do not contain y; (*b*) those which do not contain x; and (*c*) those which contain neither x nor y. Of course (*c*) is a special case of either (*a*) or (*b*).

An equation of type (*a*) is of the form $F\left(\frac{d^2y}{dx^2}, \frac{dy}{dx}, x\right) = 0$, which becomes $F\left(\frac{dy'}{dx}, y', x\right) = 0$, a first order equation in y'.

An equation of type (*b*) is of the form $F\left(\frac{d^2y}{dx^2}, \frac{dy}{dx}, y\right) = 0$, or $F\left(\frac{dy'}{dx}, y', y\right) = 0$, from which the x disappears completely when $\frac{dy'}{dx}$ is replaced by $y'\frac{dy'}{dy}$. It then takes the form of a first order equation in which the dependent and independent variables are y' and y, respectively.

To illustrate this latter type, consider the equation

$$\frac{d^2y}{dx^2} + 2\frac{dy}{dx} + y = 0, \tag{90}$$

which may be written

$$y'\frac{dy'}{dy} + 2y' + y = 0.$$

This is a homogeneous equation, and therefore y' is to be replaced by $w = \frac{y'}{y}$. This gives

$$\frac{w\,dw}{(w+1)^2} + \frac{dy}{y} = 0,$$

the solution of which is

$$\log(w+1) + \frac{1}{w+1} + \log y = \alpha.$$

If w is replaced by $\frac{1}{y}\frac{dy}{dx}$ in this equation, a new first order equation results, the solution of which is the desired relation between x and y. Unfortunately, the equation is not in a form which is amenable to the methods used above. If, however,

we are sufficiently acute to notice that w is the x-derivative of $\log y$, the substitution of a new independent variable $v = \log y$ appears to be worth trying. This gives

$$\log\left(\frac{dv}{dx} + 1\right) + \frac{1}{\dfrac{dv}{dx} + 1} + v = \alpha, \tag{91}$$

which is solvable for v. Solving and differentiating, after the fashion of § 32, it is found that

$$-\frac{dv'}{dx} = (1 + v')^2.$$

The solution of this is

$$v' + 1 = \frac{1}{\beta + x}.$$

Substituting this in (91), we get

$$-\log(\beta + x) + (\beta + x) + v = \alpha,$$

or, remembering that $v = \log y$,

$$\frac{\beta + x}{y} = e^{\beta - \alpha + x}.$$

If we replace α and β by two new constants $\alpha' = \beta\, e^{\alpha - \beta}$ and $\beta' = e^{\alpha - \beta}$, we can obtain the somewhat simpler form

$$y = (\alpha' + \beta' x)e^{-x}.$$

It can easily be verified that this is the solution of the given equation. The method used in obtaining it, however, is much more complicated than need be, as we shall see in § 62.

GENERAL PROBLEMS

1. Solve equation (22).

2. Solve equation (39).

3. Solve equation (41).

4. Solve equation (50), assuming that $v_0 = 0$, that the cathode is at $x = 0$, and that both the potential and the potential gradient are zero at the cathode.

5. Find the general solution of equation (50).

6. A sound wave is moving in the direction x. It is known that this disturbance is periodic in time, and does not vary with y or z, so that the velocity potential must take the form

$$\phi = f(x) \sin pt.$$

Find the form of the function $f(x)$. (Refer to Example 5, § 19.)

7. Solve the equation derived in Problem 5, § 20.

8. Solve (68) and (69). Verify the fact that the substitution, by means of which the latter was obtained from the former, does make these solutions identical.

9. Solve the differential equations:

$$(a)\quad \frac{dv}{du} + \frac{2v}{u} = 3v.$$

$$(b)\quad \sqrt{1 - u^2}\, dv = 2u\sqrt{1 - v^2}\, du.$$

$$(c)\quad \sqrt{1 + \frac{dv}{du}} = \frac{e^u}{2}.$$

$$(d)\quad \frac{dy}{x\,dx} = y \sin (x^2 - 1) - \frac{2y}{\sqrt{x}}.$$

$$(e)\quad \frac{dy}{dx} = 1 + \frac{2y}{x - y}.$$

10. Solve the following differential equations:

$$(a)\quad \frac{dv}{du} + 2uv = 2u.$$

$$(b)\quad (1 + v^2)\, du + (1 + u^2)v\, dv = 0.$$

$$(c)\quad u \log u \frac{dv}{du} + \sin^2 v = 1.$$

11. Solve the differential equation found in Problem 9, § 20.

12. Solve the differential equation set up in Problem 10, § 20.

13. Solve the equation set up in Problem 11, § 20, assuming that at the time $t = 0$ the velocity was zero.

CHAPTER V

Singular Solutions

§ 34. *Definition of Singular Solutions*

A differential equation sometimes has solutions which are not special cases of the general solution. How these come to exist can most easily be explained by a reference to the geometrical interpretation given in Chapter II. According to that interpretation, every particular solution corresponds to a curve, and the general solution corresponds to the family of curves taken as a whole.

To say that one of these curves satisfies the differential equation is only to say that its slope at every point, together with the coordinates of the point, satisfy the equation. If the family of curves possesses an envelope, all these quantities — slope and coordinates — are the same for a point of the envelope as for the curve to which it is tangent at that point. Hence the envelope also satisfies the differential equation. As the envelope is not in general one of the curves of the family, it cannot be obtained by assigning a particular value to the constant of integration. It therefore represents a solution that is not included in the general solution at all, and is called a *singular solution.*

These singular solutions can often be found without solving the differential equation itself: geometrically this is equivalent to saying that the envelope of the family of curves can often be found directly from its differential equation, without knowing the individual curves of the family.

In discussing envelopes in § 5, attention was called to the fact that they never exist unless more than one curve of the family pass through each point. But if several curves — say

n of them — pass through a point, each with its own appropriate slope, there must be n values of y' corresponding to each pair of values of x and y. In other words the differential equation

$$\phi(x, y, y') = 0, \tag{92}$$

when solved for y', must have n roots. This is true for any pair of values x and y. However, if the point (x, y) happens to be on the envelope of the family, one of the curves which pass through it must be counted twice, as has already been noted in § 5. Hence the value of y' which corresponds to *this* curve must appear twice as a solution of (92). The envelope of a family of curves is therefore a curve along which the differential equation of the family has equal y''s, as well as a curve along which its algebraic equation has equal c's. In the discussion of § 5 it was shown that the latter locus can be found by eliminating c between the equations

$$f(x, y, c) = 0,$$

$$\frac{\partial f}{\partial c} = 0.$$

The same argument also proves that the locus of equal y''s can be found by eliminating y' between the equations

$$\phi(x, y, y') = 0,$$

$$\frac{\partial \phi}{\partial y'} = 0.$$

The result of this elimination is an equation involving x and y only: it is not a differential equation, and it is obtained without any knowledge of the general solution of (92). That its graph includes the envelope of the family of curves defined by the differential equation is evident from what has already been said. But it may include other curves as well, as we shall see in § 35. First, however, let us illustrate the part of the theory which we have already covered.

Consider the differential equation

$$\left(\frac{dy}{dx}\right)^2 + x\left(\frac{dy}{dx}\right) = y, \tag{93}$$

the general solution of which may be found by the method of § 32 to be

$$y = cx + c^2. \tag{94}$$

This equation is linear: hence the general solution is represented by a family of straight lines. These are shown in Fig. 24. A glance at the figure shows that the lines possess an envelope C, the equation of which can be found by either of two methods:

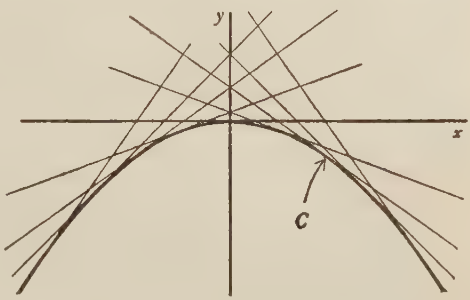

Fig. 24.

First, by using the method of § 5, which requires that (94) be differentiated with respect to c, and c then eliminated. Differentiation gives $x + 2c = 0$, which, when substituted in (94) leads to $y = -x^2/4$ as the equation of the envelope.

Second, by using the method of the present section, which requires that (93) be differentiated with respect to y', and y' then eliminated. This also leads to the same result $y = -x^2/4$.

The second of these methods could have been applied even if we had not known the solution of our differential equation.

§ 35. *Cuspidal Locus and Tac Locus*

As in § 5, so here also the locus of equal y''s contains other curves than the envelope. Usually, however, they do not represent solutions of the differential equation of the family.

If the curves which comprise the family have cusps, as in Fig. 4, the two branches which meet at the cusp have identical slopes. Hence if x and y happen to be the coordinates of a cusp, this slope will appear twice when (92) is solved for y'.

The locus of these cusps is therefore included in the locus of equal y''s, just as it was in the locus of equal c's.

It is easy to see from Fig. 4 why this cuspidal locus is not a solution of the differential equation, for the slope of the cuspidal locus is not generally the same as that of any of the family of curves passing through the point: hence, it does not satisfy the differential equation $\phi(x, y, y') = 0$, as it would do if it were a solution.

If the curves present such an appearance as that shown in Fig. 8, two equal values of y' occur at every point on the line AB at which the various curves are tangent to one another, one for each of the curves which are tangent at the point in question. Hence AB is also part of the locus of equal y''s. Like the cuspidal locus, however, it is not usually a solution of the differential equation, for the slope of AB itself is not in general the same as the slope of any of the curves. In the case of the circles, for example, their differential equation obviously requires that y' be infinite at these points, whereas the slope of AB is zero.

Such a line, at which curves touch, is called a *tac locus*.

The *nodal locus*, which was part of the locus of equal c's, is not part of the locus of equal y''s; for the two branches of the curve which cross at the double-point generally have different slopes, as in Fig. 4. Neither is it a solution of the differential equation, since the slope of the nodal locus is generally different from that of either branch.

To sum up, we have found that the locus of equal c's may consist of three types of curves: the envelope, the cuspidal locus and the nodal locus. The locus of equal y''s may also consist of three types of curves, the envelope, the cuspidal locus and the tac locus. Of these, only the envelope represents a singular solution of the differential equation.

§ 36. *Determination of Singular Solutions*

In view of the foregoing considerations it is a simple matter to obtain the singular solution of a differential equation, if it possesses any. To accomplish this, it is only necessary to find

the locus of equal c's (if the general solution is known) or of equal y''s (if it is not), and then to determine by direct substitution which factors in the equations of these loci satisfy the differential equation and which do not.

For example, consider again the family of circles shown in Fig. 8, the equation of which is

$$(x - c)^2 + y - r^2 = 0. \tag{18}$$

It has already been shown in § 5, that the locus of equal c's is composed of the straight lines $y = r$ and $y = - r$, which constitute the envelope. It has also been found in Problem 6, § 20, that the family (18) is also defined by the differential equation

$$y^2(y'^2 + 1) - r^2 = 0. \tag{95}$$

By direct substitution, the functions $y = r$ and $y = - r$ are shown to satisfy this equation: they are therefore singular solutions.

Suppose, however, that we were given the equation (95) and did not know its general solution (18). It would still be possible to determine its singular solutions directly from the differential equation itself. For upon differentiating (95) with respect to y', we would obtain

$$2y^2y' = 0, \tag{96}$$

which is satisfied by either $y' = 0$ or $y = 0$. Substituting $y' = 0$ in (95), we obtain $y = r$ and $y = - r$, which have already been seen to be singular solutions. The other function, $y = 0$, is not, since it does not satisfy the differential equation. It must, therefore, be either a cuspidal locus or a tac locus. It is, in fact, the line AB of Fig. 8, which is obviously a tac locus.

It should be noticed that $y = 0$ could only be a solution of (95) if y' were infinite. In other words, if a curve is to be a solution of (95) it must have an infinite slope wherever it intersects the line $y = 0$. That the circles have such a slope has already been noted above. It is exactly because the line $y = 0$ does *not* have this slope that it is not a singular solution.

As a second example, consider the family of curves shown in Fig. 6, the equation of which is

$$(y - c)^2 - (x + c)^3 = 0.$$

It has been shown in Problem 7, § 20, that the differential equation of the family is

$$\tfrac{8}{27} y'^3 + \tfrac{4}{9} y'^2 - y - x = 0. \tag{97}$$

Differentiation with respect to y' gives

$$y'(y' + 1) = 0,$$

the roots of which are $y' = 0$ and $y' = -1$. Substituting these values in (97) gives us the loci $y + x = 0$ and $y + x - \frac{4}{27} = 0$, respectively.

By direct substitution it is found that the function $y = -x$ is not a solution of (97). It is therefore either a cuspidal or tac locus. The function $y = -x + \frac{4}{27}$ on the other hand does satisfy (97). It is therefore a singular solution and its graph is the envelope of the curves. As $y = -x$ was also found in § 5 as a part of the locus of equal c's, it is evident that it is a cuspidal locus, as indeed Fig. 6 shows it to be.

Finally, in the case of the family of curves

$$y^2 = (x - c)(x - 2c)^2, \tag{17}$$

the differential equation has been found in Problem 8, § 20, to be

$$4yy'^3 - 2x^2y'^2 + 4xyy' + x^3 = 16y^2, \tag{98}$$

and the locus of equal y''s is easily found to be

$$(27y^2 - 2x^3)(16y^2 - x^3)^2 = 0.$$

As the locus of equal c's has already been found to be

$$y^2(27y^2 - 2x^3) = 0,$$

it follows that the line $y = 0$ and the semi-cubical parabola $16y^2 = x^3$ along which occur equal c's and equal y''s, respectively, constitute the nodal locus and the tac locus. The semi-cubical parabola, $27y^2 = 2x^3$, which is common to both, is

either an envelope or a cuspidal locus. By substitution it is found to satisfy the differential equation. The equation is therefore a singular solution, and its locus is an envelope of the curves (17). All these loci are shown in Fig. 7.

PROBLEMS

1. Obtain the locus of equal y''s for (98).

2. The equation

$$\rho\,[1 - \cos(\theta - \theta_0)] = 1$$

defines a family of parabolas in polar coordinates. Prove that the method of procedure developed in § 5 is valid in dealing with polar curves as well as Cartesian curves. Then find the envelope of the family.

3. Find a differential equation for the family of Problem 2. Find the envelope from this differential equation.

CHAPTER VI

Practical Applications of Differential Equations

§ 37. *Introduction*

In discussing the origin of differential equations in Chapter III several scientific problems were mentioned, the mathematical formulation of which led at once to differential equations. No attempt was made to solve these equations, however, as methods of solution had not then been derived. From time to time in Chapter IV these equations were again referred to and solutions were obtained for such cases as were within the power of the methods there considered. It is the purpose of the present chapter to introduce a number of additional problems of this type and to show how their solutions may easily be obtained by the theory of differential equations.

For the most part the examples chosen are extremely simple, both because the consideration of half a dozen simple examples is frequently of more value to the student than the solution of one hard problem, and also because the more difficult problems, if they are to be of any practical significance, require the explanation of so much geometry, mechanics, or electricity as to consume a large amount of time that had better be devoted to mathematical ideas. The principal aim has been to make the problems as varied as possible, and thus to give a proper perspective of the breadth of the field of application of differential equations.

§ 38. *Dissipation of Heat in a Wire*

Three fundamental laws in the theory of heat conduction were stated in § 18. For the purpose of the present example a fourth is required. It is:

(iv) A hot body surrounded by such a medium as air loses heat at a rate proportional to the difference in temperature between the two media.

The laws (i), (ii) and (iii) of § 18 are exact physical laws. That is, they are obeyed to a high degree of accuracy within very wide limits. On the other hand (iv) is only a rough approximation under any circumstances, and ceases to deserve even that rating when the difference in temperature between the body and the medium surrounding it is very great. Nevertheless, it is frequently used in the mathematical treatment of heat problems for two reasons: first, because the true law is known only in the form of empirical data; second, because the use of such data gives rise to differential equations of extreme difficulty, whereas (iv) leads to problems that can readily be solved. As long as the temperature differences amount to only a few degrees Centigrade — as in most cases of thermometry for example — the results obtained by the simple law are sufficiently accurate for most practical purposes. The following problems are supposed to be of this type.

Consider a very long wire, one end of which is maintained at constant temperature θ_0 while the rest of it is immersed in air at the temperature 0. Obviously the wire will tend to transfer its heat to the air and therefore to cool down. As it cools, heat will flow from the hotter to the cooler portions of the wire. Therefore, saying that the one end is maintained at a constant temperature implies that heat is being continually supplied at this end to make good the losses to the air.

No matter what the temperature distribution may originally have been it will ultimately settle down to a permanent state. Thereafter each point will have its own particular temperature which no longer varies with time. The problem to be solved is that of finding this "stationary temperature" at each point of the wire.

Denote by x the distance of any point from the hot end of the wire and by θ its temperature. Then the entire differential element between x and $x + dx$ will be at substantially this same temperature θ. It must, therefore, lose per unit time the

quantity of heat $g\theta\, dA$, dA being the area exposed to the air, and g the constant of proportionality in law (iv).

Also due to the temperature gradient $\frac{d\theta}{dx}$ at the point x, an amount of heat $-\, ka\, dt \left.\frac{d\theta}{dx}\right|_x$ is gained by conduction across its hotter end, a being the cross-sectional area of the wire. Similarly the amount lost across its cooler end is $-\, ka\, dt \left.\frac{d\theta}{dx}\right|_{x+dx}$. Since the temperature is stationary — that is, since it does not change with time — the gains must equal the losses, thus giving the equation

$$-\, ka\, dt \left.\frac{d\theta}{dx}\right|_x = -\, ka\, dt \left.\frac{d\theta}{dx}\right|_{x+dx} + g\,\theta(x)\, dA\, dt. \qquad (99)$$

As dx is to be infinitesimal,

$$\left.\frac{d\theta}{dx}\right|_{x+dx} = \left.\frac{d\theta}{dx}\right|_x + \left.\frac{d^2\theta}{dx^2}\right|_x dx.$$

Also dA is equal to the circumference of the wire, C, multiplied by the length of the element dx. Substituting these values in (99) and making a few simple cancellations it reduces to the form

$$\frac{d^2\theta}{dx^2} - p^2\theta = 0, \qquad (100)$$

where p^2 has been written for the constant gC/ka.

This equation is easily solved by the method explained in § 33,[1] and leads to the result

$$\theta(x) = \alpha_1\, e^{+px} + \alpha_2\, e^{-px}, \qquad (101)$$

α_1 and α_2 being the arbitrary constants of integration.

This is the general solution of the differential equation (100). The answer to our problem, however, is not a *general* solution

[1] Chapter VII contains methods better adapted for the solution of (100) than is the method of § 33; but since neither solution is difficult, we need not anticipate the more powerful method in this place.

but a *particular* one; for there is nothing arbitrary about a physical wire one end of which is maintained at a definite temperature, and there can therefore be nothing arbitrary in the mathematical description of its temperature. We must, then, seek for the boundary conditions.

One boundary condition is obvious; it is that θ must equal θ_0 when x is 0. The other condition concerns the state of the wire at large distances from the hot end. In the physical state of affairs it is obvious that this must have much the same temperature as the air surrounding it, which means that θ must approach zero for large values of x.

By inspecting (101) it is seen that the term with the negative exponent approaches zero *no matter what* α_2 *may be*, while the other approaches ∞ *unless* α_1 *is zero*. Hence the physical requirements can only be satisfied by making α_1 zero.

This fact having been established α_2 is easily determined, the final solution being $\theta(x) = \theta_0 e^{-px}$.

§ 39. *Flow of Heat in a Sphere*

As a second example, consider the problem of the stationary state of temperature within the walls of a spherical shell immersed in a constant temperature bath and containing at its center a source (such as an electric heating element) which liberates a constant amount of heat per unit time. It is obvious from the symmetry of the problem that all points equally distant from the center will have like temperatures; that is, θ is a function of r only.

Consider the element of volume contained between two imaginary concentric spheres the radii of which are r and $r + dr$. Across the inner one the heat will flow into the element at the rate $-4\pi r^2 a \frac{d\theta}{dr}$. Across the outer one heat will flow outward at a rate given by exactly the same formula, except that r^2 and $\frac{d\theta}{dr}$ must now be given the values corresponding to the outer surface. The former of these is obviously $(r + dr)^2$,

which is approximately equal to $r^2 + 2r\,dr$ if dr is sufficiently small. The latter is

$$\left.\frac{d\theta}{dr}\right|_{r+dr} = \left.\frac{d\theta}{dr}\right|_r + \left.\frac{d^2\theta}{dr^2}\right|_r dr.$$

As no heat is generated within the element of volume the temperature can only be constant provided the amount flowing in across one face balances the amount flowing out across the other. Equating these two quantities, making certain obvious cancellations, and dropping the terms that involve $(dr)^2$, the differential equation

$$\frac{d^2\theta}{dr^2} + \frac{2}{r}\frac{d\theta}{dr} = 0$$

is obtained, the general solution of which is

$$\theta = -\frac{\alpha}{r} + \beta. \tag{102}$$

The next step in the problem is to determine α and β so that (102) may satisfy the specific physical conditions laid down in the problem: first, that heat is generated at a constant rate, which may be called S units per second; second, that the outer surface of the sphere is maintained at a constant temperature, which may be called 0.

Obviously, if the temperature is stationary the rate at which heat is generated must equal the rate at which it is passed outward through any spherical shell. This latter has already been found to be $-4\pi r^2 a \frac{d\theta}{dr}$. By differentiating (102) this is found to be $-4\pi a\alpha$. Hence α must satisfy the equation $-4\pi a\alpha = S$.

At the outer surface of the sphere, where $r = R$, the temperature is zero. By substituting these values in (102), β is found to be α/R.

Therefore the final solution of the problem is

$$\theta = \frac{S}{4\pi a}\left(\frac{1}{r} - \frac{1}{R}\right).$$

It is interesting to note that the inner radius of the spherical shell does not appear in the solution of this problem. That is, the temperature distribution within a thin shell will be just the same as that within the corresponding part of a thicker one. This is not unreasonable, however, for the temperature gradient at similar points of either must be just sufficient to pass the heat outward as fast as it is generated. Hence, if the outside temperature is the same in both, it follows at once that the temperature at a given distance from the center must be the same in each.

PROBLEMS

1. A power cable, consisting of a metallic conductor five millimeters in diameter surrounded by a sheath of insulating material three millimeters thick, lies at the bottom of a bay where the temperature is substantially constant and equal to 4° C. Due to the electrical resistance of the conductor heat is generated in this cable at the rate of S units per centimeter per second. What is the temperature of the inner surface of the sheath?

2. An iron pipe of external radius 10 centimeters carries steam at a temperature of 105° C. A considerable length of it is imbedded in a cylinder of concrete the radius of which is 60 cm. The temperature of the medium surrounding the concrete wall may be taken to be 20° C. What is the temperature of a point of the concrete 20 centimeters away from the wall of the pipe?

(The thermal conductivity of iron is so high compared to that of concrete that the pipe may be regarded as being at the same temperature as the steam which it contains.)

3. Two long wires, one somewhat thicker than the other, have one end maintained at the temperature θ_0. Both are formed of the same material, and both lose heat to the surrounding air in accordance with law (iv). At unit distance from the hot end, which is the warmer?

4. A wire of length l has one end at temperature θ_0 and the other at temperature θ_1. It loses heat to the surrounding air in accordance with law (iv). How much heat does it dissipate per unit time?

5. A wire of length l has one end at temperature θ_0 and loses heat in accordance with law (iv) both from its cylindrical surface, and from the other end. What is the temperature of this other end?

§ 40. *Curve of Constant Curvature*

As the next illustration, let us take the simple geometric problem of determining the equation of a curve the curvature of which is the same at every point. It will be remembered that "curvature" is the rate at which the tangent to the curve rotates as the point of tangency moves along the curve. It is defined by the formula

$$\kappa = \frac{y''}{(1 + y'^2)^{3/2}}, \tag{103}$$

y' and y'' being derivatives of y with respect to x. The problem is therefore to solve the differential equation (103) upon the assumption that κ is a known constant.

Since $y'' = \dfrac{dy'}{dx}$, (103) integrates immediately into

$$\kappa x = \frac{y'}{\sqrt{1 + y'^2}} + \alpha.$$

Solving this for y', it is found that

$$y' = \frac{\kappa x - \alpha}{\sqrt{1 - (\kappa x - \alpha)^2}},$$

the integral of which is

$$y = -\frac{1}{\kappa}\sqrt{1 - (\kappa x - \alpha)^2} + \beta,$$

or

$$(\kappa y - \beta)^2 + (\kappa x - \alpha)^2 = 1. \tag{104}$$

This is immediately recognized as the equation of a circle of radius $1/\kappa$, the center of which is located at the arbitrary point $\left(\dfrac{\alpha}{\kappa}, \dfrac{\beta}{\kappa}\right)$. There is no necessity for determining the values of α and β, since the problem happens to be one in which the general solution is the thing desired.

§ 41. *Trajectories*

The next problem is likewise taken from the field of geometry. Chapter III discussed at some length the idea of one parameter families of curves; that is, families the various members of which were all defined by assigning different values to an arbitrary constant in a single equation. The heavy curves of Fig. 25 constitute such a family. The dotted curve possesses the peculiar property of intersecting every curve of the family at exactly the same angle. That is, if tangents are drawn to the dotted curve and one of the curves of the family at their point of intersection, these tangents include the same angle, no matter which curve of the family may have been chosen. Any curve possessing this particular property is called a *trajectory* of the family of curves.

There may be more than one such trajectory, and if so, they too constitute a family of curves. The problem of the present section is to determine this family of trajectories.

For this purpose it is first necessary to know the differential equation of the curves. How this is to be found, when the family of curves is defined by means of an algebraic equation, was explained in Chapter III. Suppose, then, that the differential equation of the curves of Fig. 25 has been found to be $f(x, y, y') = 0$. Through the point (x, y) passes a curve of the family. The direction angle θ of its tangent at this point is given by the law $\tan \theta = y'$. If a trajectory passes through the same point (x, y) its tangent must have a direction angle $\theta + \alpha$, and therefore the slope of the trajectory must be $\tan (\theta + \alpha)$.

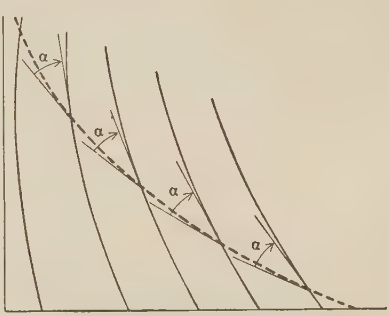

Fig. 25.

In order to avoid confusion, capital letters will be used for quantities referring to the trajectory, so that Y' is its slope at

the point (X, Y). In particular, at the point under consideration, these values must be

$$
\begin{aligned}
X &= x, \\
Y &= y, \\
Y' &= \tan(\theta + \alpha) \\
&= \frac{\tan\theta + \tan\alpha}{1 - \tan\theta\tan\alpha} \\
&= \frac{y' + \tan\alpha}{1 - y'\tan\alpha}.
\end{aligned}
$$

When solved, these equations give

$$
\left.
\begin{aligned}
x &= X, \\
y &= Y, \\
y' &= \frac{Y' - \tan\alpha}{1 + Y'\tan\alpha}.
\end{aligned}
\right\} \tag{105}
$$

Wherever a trajectory and a curve of the family cross, these relations hold. But the x, y and y' of every such point satisfy the equation (105). Hence it must be true that

$$
f\left(X, Y, \frac{Y' - \tan\alpha}{1 + Y'\tan\alpha}\right) = 0. \tag{106}
$$

This, however, is a differential equation in the capital letters: that is, it is the differential equation which the trajectories must satisfy.

As the simplest possible example, consider a family of straight lines parallel to the x-axis. The ordinary equation of such a family is $y = c$, and its differential equation $y' = 0$. But if y' is zero, Y' must equal $\tan\alpha$, as may be seen from (105). Thus the differential equation of the trajectories is $Y' = \tan\alpha$, and its solution $Y = X\tan\alpha + C$. This is the equation of a family of straight lines all of which have the

same slope $\tan \alpha$. They obviously satisfy the required condition.

As a somewhat more difficult example, let us seek to find a set of curves which cross the family of circles

$$(x - c)^2 + y^2 = c^2,$$

shown in Fig. 26, at right angles. The differential equation of this family may readily be found to be

$$y^2 - x^2 - 2xyy' = 0.$$

The family of curves crossing these circles at right angles must therefore satisfy the differential equation

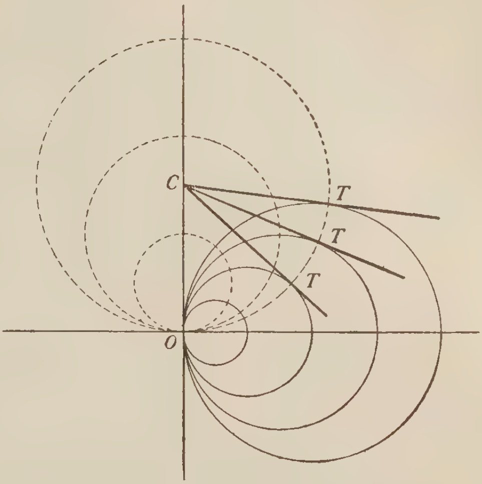

FIG. 26.

$$Y^2 - X^2 + 2\frac{XY}{Y'} = 0,$$

or

$$Y' = \frac{2XY}{X^2 - Y^2}.$$

This is a homogeneous equation the solution of which is

$$X^2 + Y^2 - 2kY = 0.$$

It represents a family of circles all of which pass through the origin and have their centers on the y-axis.

That any such circle crosses the circles of the original family at right angles can be verified from the simple geometrical fact that a line CT drawn tangent to *any* circle of the family from a point C on the Y-axis is equal in length to CO. Hence all such tangents are equal in length and may be radii of the same circle.

As a final example of this sort consider a pencil of straight

lines through the origin, the algebraic equation of which is $y = cx$, and its differential equation

$$y' = \frac{y}{x}.$$

The differential equation of trajectories crossing it at an angle α is therefore

$$Y' = \frac{X \tan \alpha + Y}{X - Y \tan \alpha}.$$

Hence the trajectories are

$$\tan^{-1} \frac{Y}{X} = \frac{\tan \alpha}{2} \log (X^2 + Y^2) + C.$$

This equation appears rather complicated when written in Cartesian coordinates. However, if polar coordinates are used, the relations

$$r^2 = X^2 + Y^2,$$
$$\theta = \tan^{-1} \frac{Y}{X}$$

reduce it at once to

$$\theta = \tan \alpha \log r + C.$$

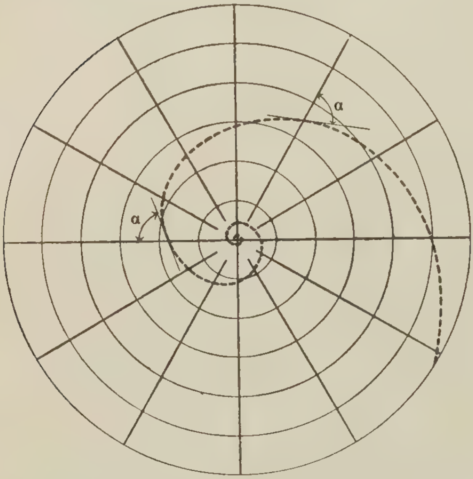

FIG. 27.—THE LOGARITHMIC SPIRAL.

These curves are called logarithmic spirals. One is shown in Fig. 27. Their most noteworthy property is that used in deriving their equations: namely, the property of crossing every radial line in the polar coordinate system at exactly the same angle α.

PROBLEMS

1. Find the equation of the trajectories crossing the polar circles $x^2 + y^2 = c^2$ at the constant angle α.

2. If the angle α is a right angle, the trajectories are called "orthog-

onal." Find the orthogonal trajectories of the family of circles $x^2 + y^2 = c^2$.

(Since polar circles and radii cross each other at right angles any trajectory of one set should likewise be a trajectory of the other, but the angles of intersection should be complementary.)

3. Find the trajectories of the family of parabolas $y^2 = cx$.

4. Find the equation of a curve the curvature of which is inversely proportional to the square of its distance from the y-axis.

5. The sine curves $y = c \sin x$ are rotated about the x-axis, thus generating a family of surfaces resembling strings of sausages. Find the equation of a set of surfaces normal to these.

§ 42. *Freely Falling Body*

It is a fundamental law of mechanics that the acceleration of a body is proportional to the force acting on it, acceleration being by definition the time rate of change of velocity. It is also a fundamental law of mechanics that the attraction exerted by one body upon another is inversely proportional to the square of the distance between them. These two laws are sufficient to determine the motion of a meteor falling upon the sun from a great distance.

Assume that at the instant $t = 0$ the meteor is at rest, that it is not impeded in its progress by frictional forces such as would be brought into play by falling through the earth's atmosphere, and that the sun is also at rest. Let the x-axis be the line joining the center of the meteor to the center of the sun, with the sun at $x = 0$ and the meteor originally at $x = X$. At any instant its velocity is $\dfrac{dx}{dt}$ and its acceleration $\dfrac{d^2x}{dt^2}$. Hence the laws stated above give the differential equation

$$\frac{d^2x}{dt^2} = -\frac{k}{x^2}. \tag{107}$$

As the acceleration is toward the sun, it tends to decrease x. This is why the negative sign is used.

The solution of (107) is found by the method of § 33. The velocity x' turns out to be

$$\frac{x'^2}{2} = C + \frac{k}{x}.$$

As the conditions of the problem require this velocity to be zero at the instant $t = 0$, when the meteor is X units from the sun, it is convenient to determine the arbitrary constant of the first integration before proceeding to the second. It is found to be $-k/X$. Hence,

$$x'^2 = 2k\left(\frac{1}{x} - \frac{1}{X}\right).$$

The solution of this equation which satisfies the condition that $x = X$ when $t = 0$ is

$$\sqrt{2kX}\; t = X\sqrt{xX - x^2} + X^2 \cos^{-1}\sqrt{\frac{x}{X}}. \qquad (108)$$

This equation gives directly the *time* at which the meteor passes a given point. It cannot easily be solved for the position x as a function of t.

The law usually given for a freely falling body is an approximation to (108) based upon the assumption that the body falls a distance which is short compared to the radius of the sun. To derive it, suppose the meteor falls to the surface of the sun from a height h above. Then x must be interpreted as the radius of the sun, and X as $x + h$. Then, if h is small compared to x,

$$\cos^{-1}\sqrt{\frac{x}{X}} = \cos^{-1}\frac{1}{\sqrt{1 + \frac{h}{x}}}$$

$$\doteq \sqrt{\frac{h}{x}};$$

this result being merely the first significant term in the Taylor's series for $\cos^{-1}\dfrac{1}{\sqrt{1 + \frac{h}{x}}}$. Also,

$$\sqrt{xX - x^2} \doteq \sqrt{hx}$$

and

$$\sqrt{2kX} \doteq \sqrt{2kx}.$$

Using these in (108) it becomes

$$h \doteq \frac{kt^2}{2x^2},$$

or, since k and x are both constant,

$$h = ct^2.$$

This law is valid so long as h is not large in comparison with the dimensions of the body upon which the meteor falls. In the case of weights let fall under the gravitational field of the earth it is always satisfactory, because the distance through which they fall is never more than a small fraction of the radius of the earth. It may be obtained more simply by assuming at the outset that the gravitational force does not vary appreciably throughout the fall.

§ 43. *Bending of a Beam*

If a beam is fixed rigidly at one end while a weight is suspended from the other, as shown in Fig. 28, the beam will bend

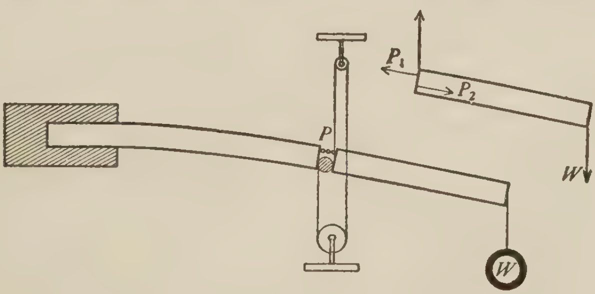

Fig. 28.—Stresses in a Bent Beam.

until the restoring forces exerted by the strained fibres of the beam are just sufficient to support the weight W. The simplest way to find the differential equation of the bent beam is to imagine the beam cut at an arbitrary point along its length and think of the forces which would have to be applied to hold

it in exactly the same position as it occupied before being cut. It is obvious, of course, that these or equivalent forces must be exerted by the particular fibres of the beam which are imagined to be cut.

Suppose such a cut were made at the point P which is located x units in a horizontal direction from the free end of the beam.[1] To support the cut portion by means of forces applied where the cut is made — and of course the cut fibres can exert forces nowhere else — requires some such arrangement as that shown in the figure.[2]

It is obvious, therefore, that

(1) the supported weight W exerts a force vertically downward;

(2) the stretched fibres in the top half of the beam exert a force P_1 in the direction of the tangent to the beam at P;

(3) the compressed fibres in the lower part of the beam exert a force P_2 which is likewise parallel to the tangent at P, but is oppositely directed to P_1;

(4) since these three forces alone could not possibly prevent the cut portion of the beam from falling, the fibres must also exert a fourth force vertically upward through the point P.

Next it must be noted that forces (2) and (3) must be of like magnitude. Otherwise there would be a resultant force in the x-direction and the cut portion would move endwise. Similarly, forces (1) and (4) must be equal, or the cut portion would move up or down. Each pair is therefore a "couple" tending only to rotate the beam. The pair (1) and (4) is called the "applied couple" and the other the "restoring couple." Since the cut portion of the beam is in equilibrium under these two couples they must have equal moments[3] — this is a funda-

---

[1] This is not the same thing as x units measured along the length of the beam, for due to deflection the beam is no longer horizontal.

[2] The rope and pulley arrangement prevents the left end of the cut portion from falling. It would not prevent the right end from falling, however. The chain and small roller accomplish this by *pushing* on the lower fibres and *pulling* on the upper ones.

[3] The moment of a force about a line is defined as the product of the magnitude of the force by the least distance between the line of action of the force and the line about which the moment is taken. In the problem considered above the line may

mental law of mechanics. As the moment of the applied couple is Wx, the restoring couple must also have this moment.

A law must now be taken from the theory of beams. It is, that the moment of the couple P_1P_2 is proportional to the curvature of the beam at the point P. Put in symbolic form, this becomes

$$\frac{qy''}{(1 + y'^2)^{3/2}} = Wx,$$

which integrates at once into

$$y = \int \frac{\frac{Wx^2}{2q} + C}{\sqrt{1 - \left(\frac{Wx^2}{2q} + C\right)^2}}\,dx + C', \tag{109}$$

in which C and C' are the constants of integration.

From the classroom standpoint this is a satisfactory solution of the differential equation, for it gives y directly as a function of x. From a practical standpoint it is not so good, for the integral term cannot be evaluated in terms of the elementary functions. In practice this difficulty is avoided by making a common-sense observation regarding the quantities that occur in the differential equation itself. Structural beams are obviously never loaded heavily enough to bend them much. Therefore $\frac{dy}{dx}$ is exceedingly small — so small in fact that the term y'^2 in the denominator may be neglected. The differential equation therefore becomes

$$qy'' = Wx,$$

and its solution

$$qy = \frac{Wx^3}{6} + Cx + C'.$$

---

be regarded as a line perpendicular to the plane of the paper through the point P. Then the moment of the force W about this line is Wx, and the moment of the vertical restoring force W' is zero, because the line of action of this force passes through the point P. The "moment of the couple" is the sum of the moments of the individual forces, and is therefore simply Wx. It will be found to be the same about any parallel line.

This solution is simple, and accurate enough to be satisfactory for many engineering purposes. It is exceedingly accurate when applied to structural members, for example. There are other branches of technology, however, in which beams occur for which the simple formula is not accurate. A watch spring, for example, is practically a bent beam, and it is certainly not true that the deflection can be ignored in such a case. To find the deflection of such a spring the complicated integral in (109) must be somehow evaluated. It belongs to a class of functions the values of which have been tabulated, just as logarithms, trigonometric functions and the Ci functions of § 25 have been tabulated. They are called "elliptic integrals," because they first attracted attention in the attempt to find the circumference of an ellipse. It is beyond the scope of this text to discuss them.

§ 44. *Deflection of Structural Columns*

The average man distinguishes a beam from a column by saying one is horizontal while the other is upright. In making this distinction, however, he is thinking primarily in terms of gravitational loads, for "beams" and "columns," to the average man, are used to support buildings. In technology the distinction is phrased a little differently, and any stiff member subjected to forces *parallel* to its own length is called a column; while if the forces are *perpendicular* to its length it is called a beam. Thus, the connecting rods and piston rods of a steam-engine are columns although they are often substantially horizontal in position, while a flagstaff is a beam so far as the wind stresses on the flag are concerned.

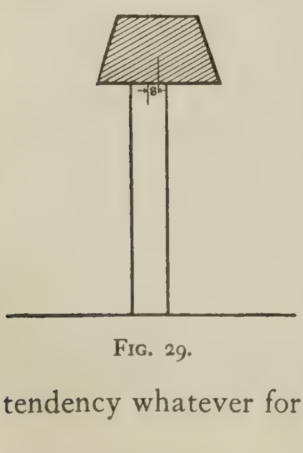

Fig. 29.

Consider the column shown in Fig. 29. If a load is applied to this column in such a way as to be distributed with perfect symmetry about its axis it is obvious that there will be no tendency whatever for the column to bend. Under these con-

ditions it will remain perfectly straight, though it will be compressed slightly. If the load is increased to such a point that the material is no longer strong enough to sustain it, it will crush without bending. In actual practice, however, no such ideal symmetry is to be expected. Instead the loads will invariably be heavier on one side than on the other, so that they are effectively equivalent to a single force applied a bit to one side of the center of the column. In the figure this distance is denoted by ϵ. Engineers speak of it as the "eccentricity of the load."

Now suppose such an eccentric load has been applied, and that the column has bent into the form shown in Fig. 30. If the column is cut at the point P it will be found that there are four forces acting upon its upper section. They are (1) the load W, (2) the tension P_1 in the outer fibres, (3) the compression P_2 in the inner fibres, and (4) the vertical force W' which prevents the cut portion from falling. As in § 43 the pair (1) and (4) form the "applied couple," and the pair (2) and (3) the "restoring couple," and to these couples we must apply the same laws that were used in discussing beams. They lead to the differential equation

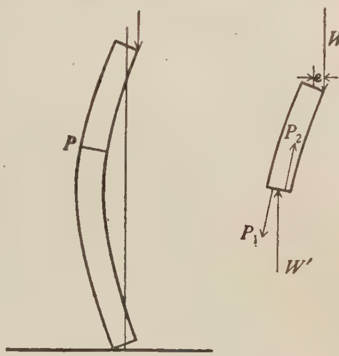

Fig. 30.—Stresses in a Bent Column.

$$\frac{qy''}{(1 + y'^2)^{3/2}} = - W(y + \epsilon), \tag{110}$$

x in this case being measured upward, and y horizontally.

Generally the deflections are small, so that y'^2 can be ignored. The equation then becomes

$$qy'' = - W(y + \epsilon),$$

the solution of which is

$$y + \epsilon = C \sin \sqrt{\frac{W}{q}}\,(x - C').$$

This is the formula usually adopted for the bending of columns; but if the deflections are large it is necessary to solve (110) without simplification.

PROBLEMS

1. A beam has a load uniformly distributed throughout its length. Find the equation for its deflection upon the assumption that the deflection is very small.

2. The beam of a chemist's balance rests upon a knife edge in the exact center and has weight pans suspended from its ends. Find the equation for its curve of deflection.

(It is simpler to think of it as being *supported* at the ends and deflected by the upward thrust of the knife edge.)

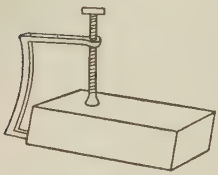

Fig. 31.

3. Find the general solution of (110).

4. The clamp shown in Fig. 31 is screwed down so tightly that its back begins to bend. What is the equation of its deflection? Assume the top and bottom arms to be absolutely rigid.

§ 45. *Vibrating String*

The problems of the last two sections concerned themselves with stiff materials, that is with materials that resist bending. A thread or a piece of soft leather obviously does not belong in this class, nor does a fine wire or a piece of gold-leaf. Such materials resist deformation only when in a stretched condition, and then not because of a reluctance to bend, but because the forces acting upon them are thrown out of equilibrium by the deformation.

As a simple example, consider a stretched violin string fastened at two points A and B. In its normal condition this string takes the form of a straight line connecting the two points. If it is deformed, its tension tends to return it to its original position, thus putting it in motion.

Let us suppose that some such deformation has taken place, and that as a consequence the string is in motion. Let us further suppose that at the time t its shape is represented by

the curve $y = f(x, t)$, of which Fig. 32 may be regarded as an exaggerated picture.[1] Finally, let us denote its tension by T. The element ds is then acted upon by two forces T_1 and T_2, each of magnitude T, but directed in slightly different directions owing to the curvature of the element. If

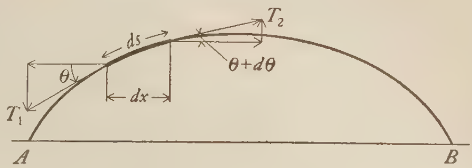

FIG. 32.—STRESSES IN A STRETCHED STRING.

the inclination of the tangent at x is denoted by θ, and that at $x + dx$ by $\theta + d\theta$, the x-components of T_1 and T_2 are $- T\cos\theta$ and $T\cos(\theta + d\theta)$, while the y-components are $- T\sin\theta$ and $T\sin(\theta + d\theta)$. If dx is exceedingly small the first two of these forces add up to[2] $- T\sin\theta\, d\theta$ and the last two to $+ T\cos\theta\, d\theta$. These forces produce accelerations of the element in the x- and y-directions, and therefore lead to the equations

$$\left.\begin{aligned} m\,ds\,\frac{\partial^2 x}{\partial t^2} &= - T\sin\theta\, d\theta, \\ m\,ds\,\frac{\partial^2 y}{\partial t^2} &= + T\cos\theta\, d\theta, \end{aligned}\right\} \qquad (111)$$

$m\,ds$ being the mass of the element considered.

If y' is very small, as it will usually be, $\cos\theta$ is nearly equal to unity while both $\sin\theta$ and θ are approximately equal to y'. Moreover, ds differs from dx by an infinitesimal of higher order than the first. Hence (111) becomes

$$\left.\begin{aligned} m\,dx\,\frac{\partial^2 x}{\partial t^2} &= - T y'\, dy', \\ m\,dx\,\frac{\partial^2 y}{\partial t^2} &= + T\, dy'. \end{aligned}\right\} \qquad (112)$$

Since y' is very small, the right-hand side of the first of

[1] We write $f(x, t)$ instead of $f(x)$ because, if the string is in motion, its shape will differ from instant to instant.

[2] We discard infinitesimals of order higher than the first.

these is nearly zero. Hence the x-component of motion may be ignored. The only important equation is therefore the second, which takes the form

$$m\frac{\partial^2 y}{\partial t^2} = T\frac{\partial^2 y}{\partial x^2}. \tag{113}$$

This is a *partial* differential equation, and therefore its comprehensive treatment is beyond the scope of this text. However, we can make use of it to answer certain simple questions by means of the tools already available. For one thing, if it may be assumed that the string vibrates back and forth in such a way that when its middle is half-way out to its maximum displacement every other point is also half-way out, and that when the middle is a third of the way out every other point is a third of the way out, and so on, we can find the frequency of its vibrations. It may be, of course, that no string would ever vibrate in this fashion. If not, the assumption that it does must lead to some absurdity which will be evidence of that fact. Either way something will have been learned.

Suppose, then, that $y = f(x)$ is the equation of the curve of maximum displacement. Obviously f is not a function of time. At any time t the string must be displaced according to the law $y = kf(x)$, where k does not vary with x, though it is a function of t. By ordinary differentiation

$$\frac{\partial^2 y}{\partial t^2} = f\frac{d^2 k}{dt^2}$$

and

$$\frac{\partial^2 y}{\partial x^2} = k\frac{d^2 f}{dx^2},$$

there being no need for round ∂'s on the right-hand side since both k and f are functions of *one* variable only. Inserting these in (113) it becomes

$$\frac{m}{k}\frac{d^2 k}{dt^2} = \frac{T}{f}\frac{d^2 f}{dx^2}. \tag{114}$$

Since f is not a function of t the right-hand side of this

equation cannot change with time; and if the right-hand side cannot change with time the left-hand side cannot either. Furthermore, since k is not a function of x, the left-hand side of the equation cannot vary with x; and if it cannot vary with x the right-hand side cannot either. Taken together, these two conclusions assert that the terms of (114) vary with neither t nor x. They are therefore constant. Their value is unknown, but may be denoted by a symbol λ, in which case (114) reduces to *two* equations

$$m\frac{d^2k}{dt^2} = \lambda k \tag{115}$$

and

$$T\frac{d^2f}{dx^2} = \lambda f. \tag{116}$$

The second of these equations is of particular interest at this stage of our argument, since it must be satisfied by the function $f(x)$, which has so far been supposed to be arbitrarily chosen. What this means is, that unless f is so chosen as to satisfy (116) the string cannot be caused to vibrate "as a whole" in the way assumed above. In other words, (116) is the answer to the question of whether or not a motion of this type is possible, the answer being that the motion is possible if and only if the deformation satisfies (116).

The solution of these two equations is not difficult, but it makes a great deal of difference in the solution whether λ is positive or negative. If we were to try both assumptions in turn, we would find that the important case is that in which λ is negative. It is then found that f must be a sine function of x, and that k is likewise a sine function of the time. The frequency of this latter sine function, which is what we set out to obtain, is $\frac{1}{2\pi}\sqrt{-\frac{\lambda}{T}}$.

In a later section this solution will be discussed at greater length. For the present the important thing is the method of solution, which depends upon the assumption that the string vibrates "as a whole"; that is, that the deflection at any

instant is obtainable from a single curve by multiplying all its ordinates by *the same* number, which changes with the time. Such a mode of motion is called a "characteristic vibration" of the string. As characteristic vibrations are of great importance in the solution of many very difficult problems, a second example of their use may be of value.

§ 46. *Vibrating Drumhead*

Suppose a drumhead is so stretched that the tension is uniform in all directions, and that it is displaced from the xy-plane into one of its characteristic modes. If it is then released, how does it behave?

Suppose to begin with that it is held in a position which

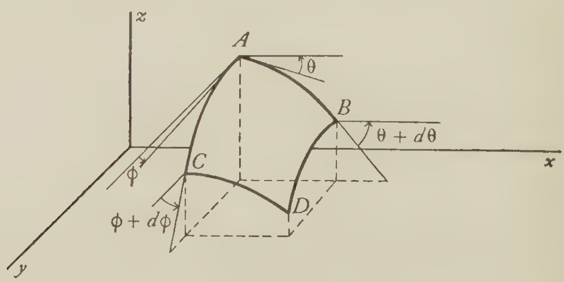

FIG. 33. — STRESSES IN A DRUMHEAD.

for the time being need not be a characteristic mode, and consider the forces acting upon a small element $dx\,dy$, Fig. 33. Across the edge AC the tension T acts in the direction θ. It gives a force $T\,dy \sin\theta$ tending to decrease [1] the displacement z. Across the opposite edge BD there is an opposing force of magnitude

$$T\,dy \sin(\theta + d\theta),$$

the resultant of the two being, except for differentials of higher order, a force

$$T\,dy \cos\theta\, d\theta$$

tending to increase [2] z.

[1] In the figure, θ is negative: hence the force tends to increase z.

[2] In the figure $d\theta$ is negative.

Similarly the other edges experience forces the resultant of which is

$$T\,dx \cos\phi\, d\phi.$$

There are also forces in the x- and y-directions, but if the deflections are kept small enough these may be ignored. If the mass of the element is $m\,dx\,dy$, its response to this force will obey the equation

$$m\,dx\,dy\,\frac{d^2z}{dt^2} = +\,T(\cos\theta\,d\theta\,dy + \cos\phi\,d\phi\,dx). \qquad (117)$$

Now, since θ is the angle between the x-axis and the tangent to a curve for which y is constant,

$$\tan\theta = \frac{\partial z}{\partial x},$$

or, to within a differential of higher order

$$\theta = \frac{\partial z}{\partial x}.$$

Likewise

$$\theta + d\theta = \frac{\partial z}{\partial x} + \frac{\partial^2 z}{\partial x^2}\,dx,$$

and therefore

$$d\theta = \frac{\partial^2 z}{\partial x^2}\,dx.$$

Similarly

$$\phi = \frac{\partial z}{\partial y}$$

and

$$d\phi = \frac{\partial^2 z}{\partial y^2}\,dy.$$

Substituting these values in (117), and noting that $\cos\theta$ and $\cos\phi$ are unity to the same degree of approximation, we get

$$m\,\frac{\partial^2 z}{\partial t^2} = T\left(\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}\right)$$

as the equation which governs the motion of the drumhead.

This equation can be solved for the characteristic modes of vibration by the same device as was used in § 45. Writing

$$z = k(t)\, f(x, y),$$

we get

$$\frac{m}{k}\frac{d^2k}{dt^2} = \frac{T}{f}\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right),$$

the round ∂'s on the right-hand side still being required since f is a function of both x and y. As in § 45, both members of this equation must be constant, which gives

$$m\frac{d^2k}{dt^2} = \lambda k,$$

$$T\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) = \lambda f.$$

The second of these equations must be satisfied by $f(x, y)$ if the drumhead is to "vibrate as a whole" in this form. If it is satisfied, as the statement of our problem presupposes, the first equation again shows that the vibration will be a sine function of the time.

In the theory of sound the sensation experienced by a person when his ears are subjected to a pressure wave that varies as a sine function of the time is called a "pure tone." Hence the last two sections have shown that the characteristic vibrations of either a stretched string or a stretched membrane are pure tones. Yet it is a matter of experience that a violin and a drum do not sound alike. The explanation is, that when either is "played" more than one characteristic tone is produced; and as the frequencies of these tones are differently related in the two cases, they give different "qualities" to the two instruments.

PROBLEMS

1. A string is stretched between the points $x = 0$ and $x = 1$. Find its characteristic modes of vibration.

(The points give boundary conditions by means of which two constants may be determined. There are, however, three unknown constants: λ, and the two constants of integration. It will be found that the boundary

values are impossible for certain values of λ, and determine only one constant of integration. That is, the two constants fixed by the physical conditions are not the constants of integration, but *one* of these, and the arbitrarily introduced constant λ.)

2. Sketch the various shapes in which the string may vibrate. What is the physical significance of the three arbitrary constants? What is the common-sense reason for expecting not to be able to fix a value for the particular one which remains undetermined?

3. Find the time-function associated with each of these characteristic modes. With what frequency does the string return to any one position?

4. What simple relation exists among the frequencies corresponding to the various characteristic modes?

5. The general time-function of Problem 3 has two arbitrary constants, and there is one remaining in the space-function of Problem 1. However, when the two functions are multiplied these combine in such a way that there are essentially only two. What physical conditions do these correspond to?

6. Assume that a circular drumhead can vibrate freely in such a way that each circle concentric with its circumference moves up and down as a whole. Find the differential equation to be satisfied by the characteristic modes of vibration. Do not attempt to solve it.

7. Solve (115) and (116) upon the assumption that λ is positive, and show that the solution does not correspond to the case of a vibrating string.

§ 47. *Surface of Revolution having Minimum Area* [1]

If two points, A and B, Fig. 34, are connected by a curve $y = f(x)$ and the whole figure is then revolved about the x-axis,

---

[1] In this and subsequent problems to which the methods of the Calculus of Variations are applied we shall make certain tacit assumptions regarding the nature of our solution the attempt to adequately explain which would carry us too far from our main purpose. The principal one, perhaps, is that the curve for which we seek is continuous and has no sharp corners. Others place restrictions upon the variations $\epsilon(x)$ which we are allowed to use. All of them are implied by the processes by means of which (120) is derived from (118); for example, when we speak of $\epsilon'(x)$ we imply at once that $\epsilon(x)$ has a derivative.

So far as our problems are concerned these restrictions are of purely theoretical interest, but there are other problems for which this is not true. For example, if we were to set ourselves the problem of finding the shape of a projectile which would be

the curve generates a surface of revolution. The area of this surface depends upon the shape of the curve drawn between the two points: that is, upon the form of the function $f(x)$.

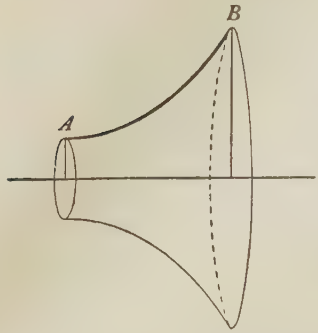

FIG. 34.

There is one curve possessing the property that the surface which it generates is of less area than that generated by any other curve. The problem is to find the equation of this minimizing curve.

As the problem is similar to those commonly given in the Calculus, in which the maximum and minimum points of curves are demanded, it will be well to review the argument by means of which such problems are solved. In the main it consists of three steps:

(i) The abscissa of the minimum point is first *assumed to be known* and represented by some letter X.

(ii) It is then noted that *a departure from this minimum in either direction must increase the function*; that is, both $f(X + \epsilon)$ and $f(X - \epsilon)$ must be greater than $f(X)$.

(iii) If ϵ is very small $f(X + \epsilon) \doteq f(X) + \epsilon f'(X)$ and $f(X - \epsilon) \doteq f(X) - \epsilon f'(X)$.[1] *Unless $f'(X)$ vanishes* one of these is greater than $f(X)$ and the other less, which violates the conclusion reached in (ii). Therefore it follows that at the minimum point the first derivative of the function must vanish.

Of course, there is more to the subject than this. It will be remembered, for instance, that the condition (iii) is necessary for a maximum point also, and it is impossible to be sure which has been obtained until the second derivative is investigated. However, this is all that is necessary for our immediate purposes.

---

offered the least possible resistance in its passage through the air, and if we were to overlook the limitations to which our methods are subject, we would not get the desired result; for it is known that the body of revolution in question is generated, not by a *smooth* curve, but by one having an abrupt corner.

[1] The symbol $\doteq$ is used in the sense of "is approximately equal to".

We shall solve our present problem by an argument which parallels these three steps exactly. In outline it is:

(i) To assume that the correct *curve* is known and has the equation $y = f(x)$.

(ii) Then if its shape is varied in any way the area of the surface of revolution *must be increased*. If the difference between the ordinates of the new curve and the old are denoted by $\epsilon(x)$ the new equation is

$$y = f(x) + \epsilon(x).$$

(iii) It can then be shown that *unless a certain differential expression vanishes*, the area generated by $f(x) + \epsilon(x)$ is greater than that generated by $f(x)$, while the area generated by $f(x) - \epsilon(x)$ is less. Hence the differential expression must vanish. This gives rise to a differential equation, the solution of which defines the desired curve.

Having thus outlined the problem, the details of part (iii) may be undertaken. Naturally, the first step is to write an expression for the area of the surface of revolution. This is a simple problem in calculus, the answer to which is

$$A = 2\pi \int_{x_0}^{x_1} f\sqrt{1 + f'^2}\, dx.$$

Suppose, now, that we replace $y = f(x)$, which gives the minimum area, by a new curve $y = f(x) + \epsilon(x)$. When this is rotated, it gives the area

$$A + dA = 2\pi \int_{x_0}^{x_1} (f + \epsilon)\sqrt{1 + (f' + \epsilon')^2}\, dx.$$

Now if ϵ represents a *small* variation in y, and if it is wisely chosen so that ϵ' is also small,

$$\sqrt{1 + (f' + \epsilon')^2} \doteq \sqrt{1 + f'^2} + \frac{f'\epsilon'}{\sqrt{1 + f'^2}},$$

and therefore

$$dA \doteq 2\pi \int_{x_0}^{x_1} \left(\epsilon\sqrt{1 + f'^2} + \frac{\epsilon' f f'}{\sqrt{1 + f'^2}} \right) dx. \qquad (118)$$

The terms not written in (118) all contain powers of ϵ higher

than the first, and may therefore be ignored by comparison with those retained.

Unless dA is zero it changes sign when the sign of ϵ is changed. This means, of course, that the area is less along one of the new curves than along the true curve itself, which is obviously in contradiction to the assumption that the true curve gives *minimum* area. It follows that dA must vanish.

Equation (118) is, in a way, the equivalent of the expression $\epsilon f'(X)$ in the simple calculus case. There is one important difference, however. In the calculus case the letter ϵ occurred only as a multiplying factor, and it therefore followed that the product could only vanish provided the other factor vanished. This statement cannot be made of (118) in its present form. Instead it must be so altered that the ϵ' disappears. First the integral is split into two parts, one composed of the term containing ϵ, and the other of the term containing ϵ'. The first is then left unchanged while the second is integrated by parts so as to give

$$\int_{x_0}^{x_1} \frac{\epsilon' f f'}{\sqrt{1 + f'^2}}\, dx = \frac{\epsilon f f'}{\sqrt{1 + f'^2}}\bigg|_{x_0}^{x_1} - \int_{x_0}^{x_1} \epsilon \frac{d}{dx}\frac{f f'}{\sqrt{1 + f'^2}}\, dx. \qquad (119)$$

As the conditions of the problem require every curve to pass through the points A and B, $\epsilon(x)$ must necessarily vanish at both limits of integration. The first term on the right-hand side of (119) is therefore zero. Substituting the remaining term in (118) gives, as the required condition for a minimum,

$$\int_{x_0}^{x_1} \epsilon(x) \left(\sqrt{1 + f'^2} - \frac{d}{dx}\frac{f f'}{\sqrt{1 + f'^2}}\right) dx = 0. \qquad (120)$$

The integrand, like the $\epsilon f'(X)$ of the calculus case, now consists of two factors: $\epsilon(x)$, which is arbitrary; and the bracketed expression, which contains only $f(x)$ and its derivatives. Just as before, this latter factor must be zero. For suppose it were not. There would then be certain intervals between x_0 and x_1 where it was negative, and others where it

was positive. As $\epsilon(x)$ is arbitrary [1] it could be chosen positive where the differential factor was negative and negative elsewhere. Then (120) would surely be negative, and the area of the surface reduced. The conclusion is reached, therefore, that the true curve $y = f(x)$ must satisfy the differential equation

$$\sqrt{1 + f'^2} - \frac{d}{dx} \frac{ff'}{\sqrt{1 + f'^2}} = 0. \tag{121}$$

This equation is so simple that its solution may be left to the reader. It will be noticed, moreover, that it is of the *second* order, and can therefore satisfy just two boundary conditions. As the problem furnishes just two such conditions in the fixed locations of the two end-points this result is quite sensible. In fact, mathematical results have an almost annoyingly monotonous habit of being sensible.

§ 48. *The Brachistochrone*

If the two points A and B, Fig. 35, are connected by a smooth frictionless wire, the shape of which is represented by

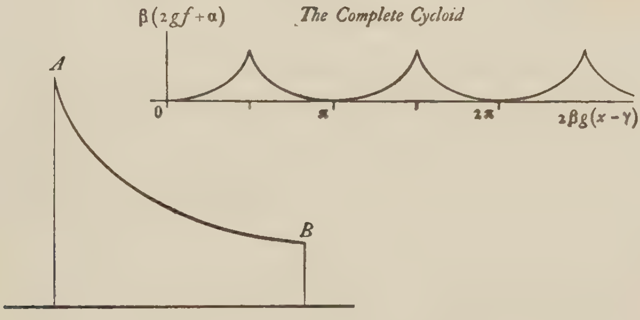

FIG. 35.—THE BRACHISTOCHRONE.

the curve $y = f(x)$, and if a weight is allowed to slide freely along this curve under the action of gravity, the time which it takes to reach B will depend upon the shape of the curve. There is, therefore, some curve along which the weight will

[1] Except that ϵ and ϵ' have been assumed to be small. Even these restrictions are not strictly necessary, but have been made to simplify the argument somewhat.

reach B in a shorter time than is required along any other curve. This curve is called the "brachistochrone" [1] between the points A and B. The problem is to find the shape of this curve.

This problem is of exactly the same type as that of § 47, but before it can be solved it is necessary to find an expression for the length of time required by the weight to slide along any wire. This can most conveniently be done by the use of three concepts from mechanics:

(i) The *potential energy* of a weight is proportional to its height above the surface of the earth. The constant of proportionality is the mass m, multiplied by the strength of gravity g.

(ii) The *kinetic energy* of a moving body is proportional to the square of its velocity. The constant of proportionality is $\frac{1}{2}m$.

(iii) The potential and kinetic energies of a body always add up to the same sum, unless it communicates energy to some other body. This is the principle of "conservation of energy."

In the problem as stated, frictional forces are absent, and therefore the weight loses no energy as it slides along the wire. Hence its kinetic energy $mv^2/2$ plus its potential energy $mg(y_0 - y)$ must always be the same. This gives the equation

$$v^2 = 2gy + \alpha, \tag{122}$$

where α is an unknown constant.[2]

Next it must be noticed that the weight is always moving in the direction of the wire. Hence v means the rate at which the arc length s is traversed; or $v = \dfrac{ds}{dt}$. Substituting this in (122) it is found that

$$\frac{dt}{ds} = \frac{1}{\sqrt{2gy + \alpha}};$$

whence the time of travel is represented by

$$t = \int \frac{ds}{\sqrt{2gy + \alpha}}.$$

[1] From two Greek words meaning "shortest time."

[2] It depends upon y_0, the height of the point A, and this can be made whatever we please by choosing the origin of coordinates properly.

Expressing ds in terms of the variable x, this becomes

$$t = \int_{x_0}^{x} \sqrt{\frac{1 + y'^2}{2gy + \alpha}}\, dx. \qquad (123)$$

This is the integral which is to be made as small as possible.

Suppose now, that $y = f(x)$ is the equation of the true curve and that $y = f(x) + \epsilon(x)$ is any neighboring curve. Along this *new* curve the time of travel may be denoted by $t + dt$ where

$$dt = \int_{x_0}^{x_1} \left(\frac{f'\epsilon'}{\sqrt{2gf + \alpha}\sqrt{1 + f'^2}} - \frac{g\epsilon\sqrt{1 + f'^2}}{(2gf + \alpha)^{3/2}} \right) dx.$$

As in § 47 it is necessary to integrate the term in ϵ' by parts, and to note that ϵ vanishes at both limits of integration. When this is done the integrand reduces to the product of two factors. As before, one of these is ϵ, which is arbitrary. Since the integral as a whole must vanish, the *other* factor must be zero, and this gives rise to the differential equation

$$\frac{d}{dx}\left(\frac{f'}{\sqrt{2gf + \alpha}\sqrt{1 + f'^2}} \right) = -\frac{g\sqrt{1 + f'^2}}{(2gf + \alpha)^{3/2}}. \qquad (124)$$

It is possible to solve this equation if the indicated differentiations are carried out, but it happens to be easier to solve it as it stands. As the process is one which is frequently useful in practical problems as well as in those of a more theoretical nature, it is worth while to carry it through step by step.

In the first place, it is noticed that the equation does not contain x. Therefore in accordance with the method of § 33, it is natural to replace the symbol $\frac{d}{dx}$ by its equivalent $f' \frac{d}{df}$. Then collecting all the f''s on the left-hand side of the equation it takes the form

$$\frac{f'}{\sqrt{1 + f'^2}} \frac{d}{df}\left(\frac{f'}{\sqrt{1 + f'^2}\sqrt{2gf + \alpha}} \right) = \frac{-g}{(2gf + \alpha)^{3/2}}.$$

On the left-hand side of this equation the quantity standing

before the symbol $\frac{d}{df}$ is almost, but not quite, equal to the one which follows it. If they were equal the entire left-hand side would be the product of a function by its derivative, and would integrate into the square of the function. It is therefore worth while to try the expedient of multiplying the entire equation by such a factor that this form will be completed. The necessary factor is obviously $\sqrt{2gf+\alpha}$ in the denominator, and the only change that it makes in the right-hand side of the equation is to replace the exponent $\frac{3}{2}$ by 2. When this change has been made the equation integrates at once into

$$\frac{f'^2}{(1+f'^2)(2gf+\alpha)} = \frac{1}{2gf+\alpha} - \beta.$$

This equation can easily be solved for f', the result being

$$f' = \sqrt{\frac{1-\beta(2gf+\alpha)}{\beta(2gf+\alpha)}},$$

whence

$$x = \int \frac{\sqrt{\beta(2gf+\alpha)}}{\sqrt{1-\beta(2gf+\alpha)}}\, df.$$

The evaluation of this integral is much simplified by changing the variable in accordance with the equation

$$\beta(2gf+\alpha) = \sin^2\theta, \tag{125}$$

in which case the integral turns out to be

$$x = \gamma + \frac{1}{2\beta g}(\theta - \sin\theta\cos\theta). \tag{126}$$

The pair of equations (125) and (126) together define the desired brachistochrone in terms of an auxiliary variable, or "parameter," θ. If we assign to this parameter a particular value, we may find a value of x from (126), and the corresponding value of $y = f(x)$ from (125). Obviously, then, by assigning one value after another to θ we could obtain as many points

upon the brachistochrone as we might desire. The curve which would result from this process is the cycloid shown in Fig. 35.

We could also eliminate θ from (125) and (126) and thus obtain the equation of the curve in its ordinary form. It is

$$\beta(2gf+\alpha)=\sin^2\left(2\beta g(x-\gamma)+\sqrt{\beta(2gf+\alpha)[1-\beta(2gf+\alpha)]}\right).$$

Instead of using this complicated equation, however, it is usually wiser to fall back upon the pair of parametric equations (125) and (126), which are simpler both to write and to compute.

§ 49. *Geodesics on a Curved Surface*

Consider two points A and B on the surface shown in Fig. 36. Among the curves which may be drawn on the surface from one of these points to the other there is one which is shorter than all the rest. It is called a *geodesic*. This geodesic is desired.

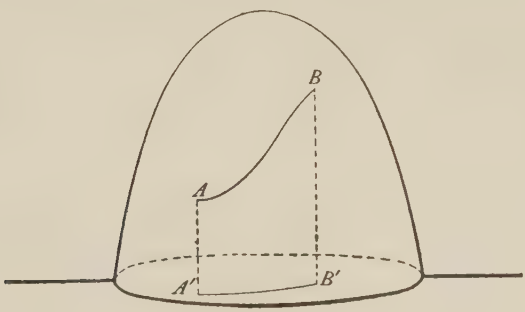

FIG. 36.

As a means of defining it, we may think of it as projected upon the xy-plane. The equation of the projected curve $A'B'$, together with that of the surface, will then fully define the geodesic.

Suppose the equation of the surface to be $z = \Phi(x, y)$. Then when x and y change by amounts dx and dy, z will change

by an amount $dz = \frac{\partial \Phi}{\partial x} dx + \frac{\partial \Phi}{\partial y} dy$. Hence the element of arc length is

$$ds^2 = dx^2 + dy^2 + dz^2 = dx^2 + dy^2 + \left(\frac{\partial \Phi}{\partial x} dx + \frac{\partial \Phi}{\partial y} dy\right)^2.$$

Now suppose the points A and B to be connected by any curve whatever. In terms of its projection $y = y(x)$ upon the xy-plane its length is

$$s = \int_{x_0}^{x_1} \frac{ds}{dx} dx = \int_{x_0}^{x_1} \sqrt{1 + y'^2 + \left(\frac{\partial \Phi}{\partial x} + \frac{\partial \Phi}{\partial y} y'\right)^2}\, dx. \qquad (127)$$

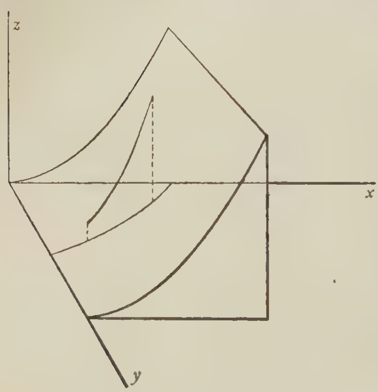

FIG. 37.

It is this integral which must be made as small as possible.

As an example take the case of the parabolic cylinder shown in Fig. 37, the equation of which is $z = bx^2$. Here $\frac{\partial \Phi}{\partial x} = 2bx$ and $\frac{\partial \Phi}{\partial y} = 0$, so that (127) becomes

$$s = \int_{x_0}^{x_1} \sqrt{1 + 4b^2x^2 + y'^2}\, dx.$$

We may obtain a differential equation from this integral by the usual process, the details of which are now sufficiently familiar that we need not repeat them. It appears in the form

$$\frac{d}{dx} \frac{y'}{\sqrt{1 + 4b^2x^2 + y'^2}} = 0. \qquad (128)$$

It is a simple matter to solve this equation. The solution, of course, gives a family of curves, any one of which possesses the property that, if two points are marked off upon it, the distance between these points is less than it would be along any other curve that could possibly be drawn between them. If we seek to find the geodesic joining two particular points, as the statement of our problem presupposes, we need only

use the coordinates of these points as boundary values and thus determine the constants of integration in our general solution.

§ 50. *The Problem of Dido*

The problems of the last few sections all require the determination of a curve along which something is as large (or as small) as possible. The study of such problems is called the *Calculus of Variations*, and is itself an important branch of mathematics. The problems so far given have belonged to the simplest type with which this study deals. That is, the curves have not been required to satisfy other conditions than that of making the integral a maximum. For example, in § 47, the only restriction placed upon the curve was that it should lead to a surface having the least possible area. Sometimes, however, it is necessary to require the curve to satisfy other conditions as well. For example, its length might be given in advance, the problem being to bend a wire of fixed length so that it would generate the minimum area. It is quite obvious that this problem has a solution, and it is equally obvious that the solution need not be that previously obtained; for it would be a remarkable coincidence if the wire furnished were of exactly the length demanded by the curve of § 47.

There are many problems of this sort. As a class they are known as *isoperimetric* problems. The one which is of greatest historical interest is called the Problem of Dido. It is said that Dido, Queen of Carthage, being in disfavor with her brother Pygmaleon, took all the money she could find and ran away to the south shore of the Mediterranean. There she struck a bargain with the king, Iarbas, for as much land as could be encompassed by a bull's hide. Then, with the fine sense of fair play which is never wanting in mythology, she cut the skin into as fine thongs as possible, tied them end to end, and made them reach around the site of Carthage. With characteristic Phœnician thoroughness, she even made the ends terminate on the seashore instead of bringing them together. Later, she committed suicide in a very spectacular

fashion to avoid marrying this same Iarbas; but that part of the story is not proper material for a text on differential equations. The point is, that having conceived this brilliant scheme, she was confronted with the problem of so placing the leather thread as to enclose the most valuable possible bit of ground — which might or might not be the largest bit, according to circumstances.

The Problem of Dido is therefore the following: Given a curve (the seacoast), and knowing the value of the land (which may vary from place to place), how can a curve of given length be drawn so that the value of the area enclosed between it and the given curve shall be a maximum?

To illustrate the method by which isoperimetric problems are treated, this Problem of Dido will be solved for the simplest possible case, in which the land is supposed to have the same value everywhere and the seashore is supposed to be straight. Moreover, it will be assumed that the ends of the string are placed at two preassigned points X units apart.[1] The problem then degenerates into that of determining what curve of given length bounds the largest area.

This curve is required to satisfy *two* conditions, one as to length and one as to area enclosed. By choosing the seacoast as the x-axis, with one end of the string at the origin, these may be written [2]

$$L = \int_0^X \sqrt{1 + y'^2}\, dx, \quad A = \int_0^X y\, dx.$$

The first is to have a fixed value, while the second is to be as large as possible.

Suppose, now, that the true curve is $y = f(x)$, its length being L_0 and the enclosed area A_0. Suppose, moreover, that

---

[1] However, see Problems 8 and 9, § 51.

[2] If the points 0 and X were too close together, it might be necessary for the thread to bulge out along the coast beyond these points, and then the integral for A would no longer be correct in the form in which we have written it. We need not concern ourselves with such complications, however. Instead, we shall assume that y is a single-valued function of x.

we were to follow our usual practice and compare this true curve with some other whose equation was $y = f(x) + \epsilon(x)$, $\epsilon(x)$ being small, but otherwise unrestricted. Obviously we could no longer say that $A_0 + dA$, the new area, was smaller than A_0; for the new curve might be longer than the old, and might therefore enclose a larger area. In other words, our usual line of argument falls down, and we are forced to seek some new method of attack.

This we shall do by shifting our attention from Dido's cord of length L_0 to a new cord the length of which is $L_0 + dL$, dL being either positive or negative. Let us suppose this new cord to be placed in such a way as to bound the largest possible area, which will be either larger or smaller than A_0, according to the sign of dL. Finally, let us denote this new area by [1] $A_0 + \Delta A$ and the differential ratio $\dfrac{\Delta A}{dL}$ (or rather, the limit of this ratio as dL approaches zero) by λ. Then we can make the statement that, *if we change the length of our curve by an amount dL, the biggest area which it can then enclose will be $A_0 + \lambda\, dL$.*

Now let us return to the consideration of our *arbitrary* comparison curve $y = f(x) + \epsilon(x)$, and let us suppose that this has a length $L_0 + dL$, which may be either longer than, shorter than, or equal to L_0. Let us also denote by $A_0 + dA$ the area bounded by this new curve. Now, whatever its length may be, the new curve cannot possibly bound an area greater than $A_0 + \lambda\, dL$, for by hypothesis this is the largest area that can be bounded by a curve of length $L_0 + dL$. Hence we must conclude that

$$dA \leqq \lambda\, dL,$$

or

$$dA - \lambda\, dL \leqq 0.$$

This leads us to the theorem: *No matter what change we may make in the curve $y = f(x)$, whether that change be such as*

---

[1] We write ΔA instead of dA because we wish to retain the latter symbol for the change in area brought about by changing to an *arbitrary* curve $y = f(x) + \epsilon(x)$.

to alter its length or not, the quantity $dA - \lambda\, dL$ *is never positive.* But if $dA - \lambda\, dL$ is never positive, $A - \lambda L$ cannot be larger for the new curve than for the old one. Hence we can restate our theorem in the more significant form:

The same curve which makes A *as large as is possible for a curve of its own length, also makes the quantity* $A - \lambda L$ *as large as is possible for a curve of any length.*

To solve the problem of maximizing A with the restriction that L must equal L_0 is therefore identical with solving the problem of maximizing $A - \lambda L$ without any restrictions whatever. It is true that the correct solution of the problem will only be obtained provided we use the right value for λ; and as it seems impossible to determine λ without knowing the solution of the problem, it may seem that little has been accomplished by this argument. We shall find, however, that if we carry λ along as an unknown constant, we shall eventually find a means of evaluating it.

The integral to be maximized is therefore

$$A - \lambda L = \int_0^X (y - \lambda\sqrt{1 + y'^2})\, dx.$$

By the usual transformations this leads to the differential equation

$$\frac{d}{dx}\frac{y'}{\sqrt{1 + y'^2}} + \frac{1}{\lambda} = 0,$$

the solution of which is

$$(x - \alpha)^2 + (y - \beta)^2 = \lambda^2.$$

This is the equation of a circle of radius λ, with its center at the point (α, β). It involves three arbitrary constants, α, β and λ; but there are also three conditions by means of which they may be determined, for the curve must pass through $(0, 0)$ and $(X, 0)$, and must also be of length L. The easiest way to determine the values of the three constants is by means of geometry. It is known that the center of a circle which passes through two points A and B lies on the perpen-

dicular bisector of AB. This means that α must be $X/2$. (See Fig. 38.) As the hypotenuse and one leg of the triangle ADC are known, the length of the other leg can easily be computed. Thus β is found to be $-\sqrt{\lambda^2 - X^2/4}$. Finally L/λ is the magnitude (in radians) of the angle ACB. The angle ACD is just half as big, and its sine is equal to $X/2\lambda$. This gives the equation

$$\sin\frac{L}{2\lambda} = \frac{X}{2\lambda},$$

from which to determine λ. This equation is transcendental and cannot be solved by algebraic means. It can be solved by "cut and try," however; or it can be solved in series. For example, if L is $1.25X$, λ turns out to be $0.552X$, and therefore β is $-0.234X$. It is this circle which is shown in Fig. 38.

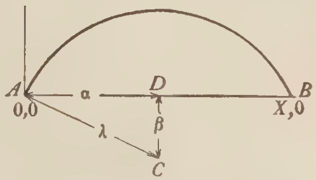

FIG. 38.

§ 51. *A Problem in a priori Probability*

It is of value to a telephone company to know how many calls to expect within a particular interval of time. Operating conditions are not always the same, and sometimes they make the solution of the problem quite difficult. But the following case is simple enough to be treated here; and besides it simulates actual conditions closely enough to have become the standard with which all other conditions are compared.

Suppose an observer is stationed in a telephone exchange to watch the calls come in, and that he counts them for a certain time interval — say a minute. What is the probability that he counts just n calls?

It will be assumed:

(*a*) That the probability that he counts just n is not influenced by what may have happened just before he started counting. That is, he may have observed that there were unusually few during several minutes while he was waiting to begin his count, or unusually many; but this, it is assumed, will not

give him any indication of what to expect during his observation. This assumption will not appear unreasonable when it is remembered that a subscriber has no means of knowing what other subscribers are doing, and therefore cannot be influenced in making his calls by what the observer has seen.

(*b*) That the chance of the observer counting just n calls is independent of the time of day when he chooses to make his test. This assumption, of course, is not true in general; for the chance of calls arriving at three o'clock in the morning is not nearly as great as at three in the afternoon. But so long as the observer is restricted to work only during those hours during which the general level of traffic is constant, the assumption is not far from right. The idea is, that whether he begins to observe at 2:35 or 2:47 or 2:51 does not matter much.

Three laws are needed from the theory of probability:

(i) If a thing is sure to happen, its probability is 1.

(ii) If either of two things can happen quite independently of whether the other does or not, the chance that *both* happen is the *product* of the probability of one happening by the probability of the other happening.

(iii) If it is impossible for both of two things to happen, the chance that *either one or the other* happens is the *sum* of the chance that the first happens plus the chance that the other happens.

We are now ready to carry out the solution of our problem. It will consist of three parts. In Part I we will obtain the chance that the observer counts *no* calls. In Part II we will derive a differential equation for his chance of counting just n calls. Then in Part III we will solve this equation.

Part I. — Call the length of time during which the observation lasts t, and divide it into a large number of very small intervals dt. Then there will be no calls in t if and only if there are none in any of these small intervals. Hence the chance of no call in t, which will be written $p_0(t)$, is the product of similar probabilities for all the small intervals. This follows from law (ii). That is:

$$p_0(t) = [p_0(dt)]^{t/dt}, \tag{129}$$

where t/dt is the number of intervals.

Next, let $q(dt)$ represent the chance that a call *does* occur in such a small interval dt. As an interval cannot at the same time have a call and not have a call, law (iii) asserts that $p_0(dt) + q(dt)$ is the probability that a call either does or does not occur in such an interval. Obviously one or the other of these things must happen: hence by the first law the sum $p_0(dt) + q(dt)$ is equal to unity. Therefore (129) becomes

$$p_0(t) = (1 - q)^{t/dt};$$

or

$$\log p_0(t) = \frac{t}{dt} \log (1 - q).$$

If dt is very small — say a thousandth of a second — the chance that it contains *one* call is very small; and the chance that it contains *two* calls or more may be neglected entirely: q then is exceedingly small. But if q is small, $\log (1 - q)$ is approximately equal to $- q$; whence

$$\log p_0(t) \doteq - \frac{tq}{dt}. \tag{130}$$

Next, let dt get smaller and smaller: evidently the left-hand side of (130) cannot be changed by this means, for it represents the chance of no calls in the entire interval t; and if the left-hand side does not vary the right-hand side cannot. This requires that q contain dt as a factor. Call it

$$q = k\,dt; \tag{131}$$

then (130) reduces to

$$p_0(t) = e^{-kt}.$$

This is the probability of no calls during the interval of observation. The k, of course, is unknown; but it is not without its physical significance, for (131) states that $k\,dt$ is the chance of a call in an interval dt, provided dt is *very* small.

Part II. — For the second part of the problem, take an interval of any length t, followed by a *very short* interval dt. Taken together they form an interval of length $t + dt$, which will be called the "combined interval."

If there are n calls in this combined interval, they must either all be in t and none in dt; or $n - 1$ in t and one in dt; or divided between them in some other way. The other ways all require more than one call in dt, and are therefore so extremely unlikely that they may be ignored.

The chance of n calls in t and none in dt is

$$\begin{aligned} p_n(t)p_0(dt) &= p_n(t)(1 - q) \\ &= p_n(t)(1 - k\,dt), \end{aligned}$$

and the chance of $n - 1$ calls in t and one in dt is

$$p_{n-1}(t)\,p_1(dt) = p_{n-1}(t)\,k\,dt.$$

Hence the chance of n calls in the combined interval is

$$p_n(t + dt) = p_n(t)(1 - k\,dt) + p_{n-1}(t)\,k\,dt,$$

which can easily be thrown into the form

$$\frac{p_n(t + dt) - p_n(t)}{dt} = k\,[p_{n-1}(t) - p_n(t)]. \tag{132}$$

This equation is only true provided dt is very small; even then it can only be regarded as a very good approximation, for the chance of more than one call in dt though small, is not absolutely zero. However, the approximation is improved as dt becomes smaller, and as dt vanishes (132) becomes exact. So though (132) as written is only approximate, the differential equation

$$\frac{dp_n}{dt} = k\,(p_{n-1} - p_n), \tag{133}$$

to which it leads, is absolutely true. It is only necessary to solve it to obtain the value of p_n.

Part III. — Suppose n to be taken equal to unity. Then (133) becomes

$$\frac{dp_1}{dt} + k\,p_1 = k\,p_0.$$

As p_0 is known to be e^{-kt}, this is an ordinary linear equation, the solution of which is

$$p_1 = (\alpha_1 + kt)e^{-kt}.$$

The determination of the proper value for the constant of integration α_1 is not difficult if the meanings of the quantities are kept in mind, for the chance of one call in a *very small* interval must be very small, and must approach zero as the length of the interval approaches zero. That is, p_1 must be zero when t is zero, and therefore α_1 must vanish. Hence

$$p_1 = kte^{-kt}.$$

Similarly, for $n = 2$ (133) becomes

$$\frac{dp_2}{dt} + kp_2 = k^2t\,e^{-kt},$$

the solution of which is

$$p_2 = \left(\alpha_2 + \frac{k^2t^2}{2}\right)e^{-kt},$$

where α_2 must likewise be zero. Proceeding in this fashion $p_3, p_4, \cdots$, can all be found in succession.[1]

---

[1] An interesting alternative method of solving (133) makes use of the following artifice:

We first introduce an auxiliary variable λ, and define a function $F(\lambda, t)$ by the Taylor's series

$$F(\lambda, t) = p_0 + \lambda p_1 + \lambda^2 p_2 + \cdots. \qquad (a)$$

About this series we note two things: first, that when $\lambda = 1$ it reduces to $F(1, t) = 1$; and second, that it must therefore be convergent for all smaller values of λ.

Next, we differentiate our series and replace the derivatives of the p's by (133). Thus we obtain

$$\frac{dF}{dt} = k(\lambda - 1)F,$$

or

$$F(\lambda, t) = \alpha(\lambda)\,e^{k(\lambda-1)t}, \qquad (b)$$

where $\alpha(\lambda)$ is the "constant of integration."

To determine this constant we first notice that when $t = 0$ all the p's except p_0 vanish. Hence (a) reduces simply to

$$F(\lambda, 0) = p_0(0).$$

Likewise (b) reduces to

$$F(\lambda, 0) = \alpha(\lambda).$$

Again a differential equation has proved to be the clue to the solution of an important technical problem.

PROBLEMS

1. Find the shortest path between the two points (x_0, y_0) and (x_1, y_1) in the xy-plane.

2. A body must be moved from the point (1,0) to the point (2,2). The nature of the surface over which it is to be moved varies from place to place in such a way that the labor expended in moving it a unit of distance is proportional to its distance from the origin. Find the path that requires the least expenditure of energy.

3. A load is to be moved from $x = 0$ to $x = 1$ in one second. There are many ways of accomplishing this: by moving it at unit velocity all the way, or with velocity 2 from $x = 0$ to $x = \frac{2}{3}$ and velocity $\frac{1}{2}$ from there on, or in general with the velocity implied in the statement $x = f(t)$. In doing this it is necessary to overcome a resisting force proportional to the velocity. What should the schedule be in order to expend the least possible energy?

(This problem can be formulated as an integral from $x = 0$ to $x = 1$, in which case it is an isoperimetric problem; or as an integral from $t = 0$ to $t = 1$, in which case it is not. See if you can get the same answer both ways. Energy is $\int F\,dx$, or $\int F \frac{dx}{dt}\,dt$, where F is force.)

4. A cubic centimeter of brass is to be worked into a solid of revolution, the axial diameter of which is one centimeter. It is desired to make its moment of inertia about the axis as small as possible. What shape should it be given?

Hence $\alpha(\lambda) = p_0(0)$, and thus is a constant. On the other hand, if we set λ equal to 1 in (b) and remember that $F(1, t) = 1$, we find that the value of this constant $\alpha(\lambda)$ is unity. That is,

$$F(\lambda, t) = e^{-kt} e^{\lambda kt}.$$

Let us now replace $e^{\lambda kt}$ by its Taylor's series. The result is,

$$F(\lambda, t) = e^{-kt} + \lambda\, kt\, e^{-kt} + \lambda^2 \frac{k^2t^2}{2!} e^{-kt} + \cdots. \qquad (c)$$

Thus again we have found a convergent Taylor's series for $F(\lambda, t)$. But it is impossible for a function to be represented by two different Taylor's series. Hence the coefficients in (a) and (c) must be equal; which gives us

$$p_0 = e^{-kt}, \quad p_1 = kt\,e^{-kt}, \quad p_2 = \frac{k^2t^2}{2!}e^{-kt},$$

as before.

(The moment of inertia of a mass m about an axis d units distant from it is md^2.)

5. A deity, about to create a world out of a given amount of matter at his disposal, wants it to satisfy three conditions: first, that it shall be a solid of revolution; second, that it shall be homogeneous; and third, that it shall exert the maximum possible gravitational attraction upon a body placed at its "north pole."

Being all-wise, the deity set up his differential equation in terms of polar coordinates, using the "north pole" as origin, and found the proper shape. I am sure the reader will be gratified to find that he can do the same, and will wish to draw a curve representing the cross-section of this new world.

6. Find the curve of length 3 joining the points $(-1, 1)$ and $(1, 1)$, which, when rotated about the x-axis, will give minimum area.

(In obtaining the solution of this problem you will find that one of your constants can take either of two values. Obviously there is only one solution to the problem. Which of the two values is the correct one, and what does the other mean?)

7. Find the surface of revolution which encloses the maximum volume in a given area.

(After obtaining a differential equation containing only y' and y, the substitution $1 + y'^2 = 1/u$ will greatly simplify the work. You will be unable to evaluate your last integral in general. If, however, you require the surface to meet the axis of rotation *at right angles* — which merely says that it must not have pointed ends — you will obtain the solution you expect. What is the significance of the other solutions?)

8. Dido was confronted also with the problem of determining how far apart the end-points A and B, Fig. 38, should be made, in order to get the most out of her bargain. Answer this question by finding the area in terms of the angle ϕ (that is, ACD) and determining the value of ϕ for which it is a maximum.

9. Land on the Carthage waterfront probably was of greater potential value than further inland. This somewhat complicated Dido's placing of the bull's hide. Suppose land values varied according to the law

$$v(\eta) = \frac{1}{1 + \eta}$$

$v(\eta)$ being the value of a unit of area located at a distance η from the sea. Find the boundary, all other conditions being as in the text. Do not attempt to perform the last integration.

CHAPTER VII

Linear Equations of Order Higher than the First

§ 52. *Introductory Remarks*

The reader will find, upon referring to the examples contained in §§ 45 and 46 and to his solution of the problems which accompanied those examples, that all of the differential equations were linear. This was not merely a coincidence. Instead, it may be stated as a general rule that the vast majority of problems dealing with elastic deformations formulate themselves as linear differential equations. This is true also of most problems in the conduction of heat or of electricity, and of many other technical and scientific problems as well. In fact, so great is the preponderance of linear equations in the more usual fields of physical research that the study of the differential equations of mathematical physics might almost equally well be called the study of linear equations.[1]

It is not out of place, then, to give special attention to equations of this form; and accordingly the remainder of this text is devoted almost entirely to their consideration.

They divide themselves sharply into two classes: those with *constant*, and those with *variable*, coefficients. Of these classes, the latter is generally (and properly) made the subject of a separate course of study. We shall therefore confine our attention in the main to the simpler case of constant coefficients, the only exceptions being the introductory remarks of §§ 53 and 54, which apply equally well to all linear equations, and certain types which are mentioned in Chapter IX for the purpose of giving the reader an intimation of a direction in which he may profitably extend his studies if he is so inclined.

[1] The reason for this preponderance of linear equations is, of course, the fact that in the vast majority of scientific problems we are content to deal with first order effects.

Throughout this study we shall find it convenient to denote the differential expression

$$f_s(x)\frac{d^s y}{dx^s} + f_{s-1}(x)\frac{d^{s-1}y}{dx^{s-1}} + \cdots + f_1(x)\frac{dy}{dx} + f_0(x)y$$

by the shorthand notation $F_s\left(x, \frac{d}{dx}\right)y$. This notation is suggested by the fact that derivatives of various orders are written as if they were powers of $\frac{d}{dx}$, so that the whole expression has the appearance of y multiplied by a function of two variables x and $\frac{d}{dx}$. Of course such a function would naturally be represented by $F_s\left(x, \frac{d}{dx}\right)$, if we wished to call specific attention to the fact that the highest power of $\frac{d}{dx}$ was the sth. In this notation the most general linear equation of the sth order is

$$F_s\left(x, \frac{d}{dx}\right)y = f(x),$$

$f(x)$ being any arbitrary function of x.

§ 53. *The Principle of Superposition*

The most important general theorem relating to linear equations is known as the *principle of superposition.* It deals with two equations,

$$F_s\left(x, \frac{d}{dx}\right)y = f(x) \tag{134}$$

and

$$F_s\left(x, \frac{d}{dx}\right)y = g(x), \tag{135}$$

which are exactly alike on the left-hand side, but have different functions of x on the right; and with a third equation,

$$F_s\left(x, \frac{d}{dx}\right)y = f(x) + g(x), \tag{136}$$

which is also identical on the left, but has as its right-hand member the sum of the right-hand members of the other two. The content of the theorem is, that the sum of a solution of (134) and a solution of (135) is a solution of (136).

It is convenient to think of all three equations as identical in the sense that they have the same differential structure, and to speak of the solutions as being "due to" the particular form of the right-hand member. The theorem then takes the form:

THEOREM I. — *The sum of solutions of a differential equation due to two functions* $f(x)$ *and* $g(x)$ *is a solution due to the function* $f(x) + g(x)$.

In order to prove this statement, denote the solutions of (134) and (135) by $\phi(x)$ and $\psi(x)$, respectively. In the language of the theorem these are the solutions "due to f" and "due to g," respectively. Then the supposed solution "due to $f + g$" is $\phi + \psi$.

To determine whether our theorem is true or false, then, it is only necessary to substitute it in (136) and see whether the equation is satisfied. However, since any derivative of $\phi + \psi$, whatever its order, is the sum of derivatives of like order of ϕ and ψ separately, when $\phi + \psi$ is substituted in (136) the left-hand member breaks up into a part containing only ϕ, and another containing only ψ. By comparison with (134) the first of these is seen to be equal to f, and by comparison with (135) the second is equal to g. Hence the theorem is established.

So far ϕ and ψ have been spoken of as if they were *particular* solutions. However, if ϕ is the *general* solution of (134) and ψ a *particular* solution of (135), $\phi + \psi$ must be the *general* solution of (136). For if ϕ is the general solution of (134) it contains s independent arbitrary constants. Therefore $\phi + \psi$ also contains s independent arbitrary constants, and the general solution of (136) can contain no more.[1]

[1] It might seem, offhand, that by adding together the *general* solutions due to f and g, an even more general solution of (136) would be obtained, because of the $2s$ arbitrary constants. But only s of these constants are independent. Hence there is no gain in generality, as indeed our remarks above should have led us to expect.

Passing to a special case, if $g = 0$, (134) and (136) are identical on *both* sides, and (135) becomes the corresponding equation with its second member suppressed, or, as it is usually called, the *complementary equation.* For this special case Theorem I may be stated as follows:

THEOREM II. — *The general solution of a linear differential equation is the sum of any one of its particular solutions and the general solution of the complementary equation*;[1] or, in other words, *The general solution of a differential equation "due to $f(x)$" is the sum of any particular solution "due to $f(x)$" and the general solution "due to zero."*

Nearly all methods of solving linear equations are built upon the principle of superposition, or upon Theorem II, which is a special case of it. The principle should therefore be thoroughly mastered in both forms. In a later section they are interpreted in terms of electrical circuit theory, and it will be found that each of them has a simple physical meaning.

§ 54. *The Principle of Decomposition*

It is to be noted that, though the solution of (136) can be built up once the solutions of (134) and (135) are known, the process cannot generally be inverted. That is, if a solution due to $f + g$ is known, it is not generally possible to state what part of it is due to f and what part to g. A similar state of affairs occurs in elementary arithmetic. The two numbers 3 and 4 being known, their sum 7 is uniquely fixed: but if it is

---

[1] The general solution of the complementary equation is usually called the "complementary solution" of the original equation. In spite of its name, however, this complementary solution is not a solution at all; for it reduces the left-hand member to zero, not to $f(x)$. For this reason the term "complementary solution" is avoided in this text, in favor of the longer phrase "solution of the complementary equation." Even this phrase is unsatisfactory, for it is difficult to see in what sense the equations are complementary; but at least it is not wholly misleading. The "due to" terminology is based upon the general dynamical significance of the equation, in which y is the displacement "due to" a force $f(x)$. Personally, I like it well enough to dispense with the other entirely, were it not for the fact that the reader might then find the term "complementary" meaningless when he met it in print.

known that the sum of two numbers is 7, the numbers cannot be inferred therefrom. They may be 3 and 4, or 2 and 5, or any other of an infinity of pairs.

There is one special case, however, in which the arithmetical process can be inverted and the sum split up into its original terms: if the sum is a complex number $a + bi$, and is known to have been obtained by adding together a *real* and an *imaginary* term, it follows that the real term must have been a, and the imaginary one bi. In all other cases where the sum of two numbers is $a + bi$, at least one is complex. Similarly there is a special case of the principle of superposition in which the solutions due to f and g can be inferred from the solution due to their sum, and it so happens that this special case is entirely analogous to the simple arithmetical example which we have just used.

Suppose $f(x)$ is real, while $g(x)$ is a pure imaginary and equal to $i\,g_1(x)$; suppose further that the coefficients of all the derivatives in (136) are real. Finally, suppose that a solution of (136) is known. This solution will have a real part and an imaginary part, and may be written $y = \phi(x) + i\,\psi_1(x)$. When this is substituted in (136), the equation may readily be reduced to the form

$$F_s\left(x, \frac{d}{dx}\right)\phi + i\,F_s\left(x, \frac{d}{dx}\right)\psi_1 = f(x) + i\,g_1(x).$$

Since ϕ and ψ_1 are *real* quantities, their x-derivatives are also real. Therefore $F_s\left(x, \frac{d}{dx}\right)\phi$ is real and $i\,F_s\left(x, \frac{d}{dx}\right)\psi_1$ imaginary. It follows, then, from the fact that two complex quantities are equal only when their real and imaginary parts are separately equal, that

$$F_s\left(x, \frac{d}{dx}\right)\phi = f(x)$$

and

$$F_s\left(x, \frac{d}{dx}\right)\psi_1 = g_1(x).$$

The first of these states that ϕ is the solution due to $f(x)$, and the second that ψ_1 is the solution due to $g_1(x)$.

In words this theorem may be stated as follows:

THEOREM III. — *If, in the linear differential equation*

$$F_s\left(x, \frac{d}{dx}\right) y = f(x), \tag{134}$$

the coefficients of the derivatives are real while $f(x)$ is complex, the real and imaginary parts of the solution due to $f(x)$ are due, respectively, to the real part of $f(x)$ and to its imaginary part.

One clause in this theorem needs special mention: If a coefficient in $F_s\left(x, \frac{d}{dx}\right)$ is *not* real, the result of replacing y by a real function of x in that term is not real. Hence it is no longer possible to infer that both $F_s\left(x, \frac{d}{dx}\right) \phi$ and $F_s\left(x, \frac{d}{dx}\right) \psi_1$ in (136) are real. This, of course, destroys the argument. The second example below is a case of this sort, and may serve to emphasize the need of requiring that *all* the coefficients be real.

Similarly, if an equation is not linear the theorem does not apply; for in such a case y or one of its derivatives must be multiplied either by itself or by another derivative. But if the y which solves (134) is complex, its derivatives are also, so that any such product is a product of two complex quantities. As it is *not* true that the real part of such a product is the product of the real parts of the factors, nor the imaginary part the product of the separate imaginaries, the argument breaks down.

As a first example, take the equation

$$\frac{dy}{dx} + \frac{2}{x} y = x + ix^2. \tag{137}$$

The solution of this equation is easily found, by the method of § 30, to be

$$y = \frac{x^2}{4} + \frac{ix^3}{5} + \frac{c}{x^2}. \tag{138}$$

Similarly, the solutions of

$$\frac{dy}{dx} + \frac{2}{x} y = x$$

and

$$\frac{dy}{dx} + \frac{2}{x} y = ix^2$$

are, respectively,

$$y = \frac{x^2}{4} + \frac{c_1}{x^2}$$

and

$$y = \frac{ix^3}{5} + \frac{ic_2}{x^2}.$$

They are therefore indeed the real and imaginary parts of (138), provided the arbitrary constant c is complex, and has the real and imaginary parts c_1 and ic_2, respectively.

If, however, (137) is replaced by

$$\frac{dy}{dx} + \frac{2i}{x} y = x + ix^2, \tag{139}$$

so that the coefficient of y is no longer real, its solution is

$$y = \frac{x^2}{2i + 2} + i \frac{x^3}{2i + 3} + \frac{c}{x^{2i}},$$

or

$$y = [\tfrac{1}{4} x^2 + \tfrac{2}{13} x^3 + a \cos (b - 2 \log x)]$$
$$+ i [- \tfrac{1}{4} x^2 + \tfrac{3}{13} x^3 + a \sin (b - 2 \log x)],[1] \tag{140}$$

[1] The separation of real and imaginary parts in the case of the first two terms is simple: it requires only multiplication by the conjugate imaginaries of the denominators. The method of separation in the last term may not be so familiar to the reader. Using the value $c = a e^{ib}$ and writing

$$x^{-2i} = e^{-2i \log x},$$

the term takes the form

$$a e^{i(b - 2 \log x)}.$$

But

$$e^{i\theta} = \cos \theta + i \sin \theta.$$

Hence, upon identifying $b - 2 \log x$ with θ, the last term becomes

$$a \cos (b - 2 \log x) + i a \sin (b - 2 \log x).$$

provided the arbitrary constant c is expressed in the form $a\,e^{ib}$. The real and imaginary parts are here separated, but they are *not* solutions of

$$\frac{dy}{dx} + \frac{2i}{x} y = x$$

and

$$\frac{dy}{dx} + \frac{2i}{x} y = ix^2.$$

Instead, the solutions of these are, respectively,

$$\left.\begin{aligned} y &= \frac{x^2}{2i + 2} + \frac{c_1}{x^{2i}}, \\ y &= \frac{ix^3}{2i + 3} + \frac{ic_2}{x^{2i}}, \end{aligned}\right\} \tag{141}$$

and are not even real.

This outcome was to be expected, since the *principle of decomposition* (Theorem III) does not apply if the coefficients of the differential terms are complex, as in (139). However, no such restriction was made in developing the *principle of superposition* (Theorem I). The sum of the terms (141) ought therefore to be equal to (140), as indeed they are.

Finally, as a third example, consider the equation

$$y\left(\frac{dy}{dx} + \frac{2i}{x} y\right) = x + ix^2 \tag{142}$$

the solution of which is easily found to be

$$y = \sqrt{\frac{x^2}{2i + 1} + \frac{2ix^3}{4i + 3} + \frac{c}{x^{4i}}}. \tag{143}$$

Likewise, the solutions of the equations

$$y\left(\frac{dy}{dx} + \frac{2i}{x} y\right) = x$$

and

$$y\left(\frac{dy}{dx} + \frac{2i}{x} y\right) = ix^2$$

are

$$y = \sqrt{\frac{x^2}{2i + 1} + \frac{c_1}{x^{4i}}} \tag{144}$$

and

$$y = \sqrt{\frac{2ix^3}{4i + 3} + \frac{ic_2}{x^{4i}}}. \tag{145}$$

But not only are (144) and (145) not the real and imaginary parts of (143): their sum is not even equal to (143). That is, in this case both the principle of decomposition and the principle of superposition fail, as indeed they ought, since (142) is not a linear equation.

§ 55. *Linear Equations with Constant Coefficients; Exponential Solutions*

The principle of superposition applies to all linear equations, no matter what the nature of the coefficients; and the principle of decomposition applies whenever the coefficients are real. But the processes which are now to be explained generally fail unless these coefficients are constant. Hence, *throughout the remainder of the chapter only those equations will be discussed which are linear and have constant coefficients.*

Any such equation is of the form

$$a_s \frac{d^s y}{dx^s} + a_{s-1} \frac{d^{s-1} y}{dx^{s-1}} + \cdot \cdot \cdot + a_1 \frac{dy}{dx} + a_0 y = f(x), \tag{146}$$

or, in symbolic notation,

$$F_s\left(\frac{d}{dx}\right) y = f(x), \tag{147}$$

which differs from (134) only in the fact that x is no longer an argument of the function F_s. As the solution of all such equations may be made to depend upon that obtained when $f(x)$ is an exponential of the form Be^{px}, their study can best be introduced by considering this special case.

Suppose, then, that the differential equation is

$$F_s\left(\frac{d}{dx}\right) y = Be^{px}. \tag{148}$$

A particular solution can then always be found in the form $y = Ae^{px}$, unless $F_s(p) = 0$. For if y has this form, its successive derivatives may be written down at once by merely replacing $\frac{d}{dx}$ by p. Thus $\frac{d}{dx}y = py$, $\frac{d^2}{dx^2}y = p^2y$, and so on. Hence when y is assumed to have this form (148) becomes

$$F_s(p)Ae^{px} = Be^{px}. \tag{149}$$

This is no longer a differential equation, for p is merely a number and therefore $F_s(p)$ is also a number. The only undetermined quantity in it is A, and the equation shows at once that A must be given the value

$$A = \frac{B}{F_s(p)}.$$

That is, a particular solution of (148) is

$$y = \frac{Be^{px}}{F_s(p)}.$$

This result can be formulated as a theorem as follows:

THEOREM IV. — *If the right-hand side of a linear equation with constant coefficients is an exponential function Be^{px}, as in (148), and if $F_s(p) \neq 0$,*

$$y = \frac{Be^{px}}{F_s(p)} \tag{150}$$

is known to be a particular solution.

As an example, consider the equation

$$\frac{d^2y}{dx^2} + 3\frac{dy}{dx} + 2y = 6e^{-5x}. \tag{151}$$

If a solution is assumed in the form

$$y = Ae^{-5x}$$

and substituted in (151), it leads to

$$12Ae^{-5x} = 6e^{-5x},$$

which is a true equation if, and only if, $A = \frac{1}{2}$. It is easily seen that

$$y = \tfrac{1}{2}e^{-5x} \tag{152}$$

is therefore indeed a solution.

As the statement of Theorem IV indicates, there is one exceptional situation in which this device for obtaining a particular solution does not succeed. When p happens to be a root of $F_s(p)$, the left-hand side of (149) is zero no matter what value A may have, and the equation cannot be satisfied. The trouble is, of course, that when p is a root of $F_s(p)$, (148) does not have a solution of the form assumed. In § 63 a method will be developed by means of which the correct solution may be obtained even in this exceptional case.

The very property which makes the case exceptional, however, is of immense importance in another sense. For the statement "When p is a root of $F_s(p)$, the left-hand side of (149) is zero no matter what value A may have" is equivalent to the statement "If p is a root of $F_s(p)$, Ae^{px} is a solution of the equation complementary to (148), no matter what the value of A may be"; and by means of this observation we may easily get the *general* solution of the complementary equation. For the equation $F_s(p) = 0$ is an ordinary algebraic equation of the sth degree,[1] and therefore has s roots. If we denote these roots by $p_1, p_2, \cdots, p_s$, *every one of the functions*

$$\begin{aligned} y &= \alpha_1 e^{p_1 x}, \\ y &= \alpha_2 e^{p_2 x}, \\ &\cdots \quad \cdots \\ y &= \alpha_s e^{p_s x} \end{aligned}$$

is a solution of the complementary equation, no matter what the values of $\alpha_1, \alpha_2, \cdots, \alpha_s$ *may be.*

---

[1] It is often called the *auxiliary* equation of the differential equation (147). The reader will note that the *complementary* equation is a differential equation; while the *auxiliary* equation is algebraic.

Moreover, by the principle of superposition their sum

$$y = \alpha_1 e^{p_1 x} + \alpha_2 e^{p_2 x} + \cdots + \alpha_s e^{p_s x} \qquad (153)$$
$$= \sum_{j=1}^{s} \alpha_j e^{p_j x}$$

is also a solution; and since it has s arbitrary constants it must be the *general* solution, unless they happen not to be independent. A moment's consideration serves to show that, in case any two of the numbers $p_1, p_2, \cdots, p_s$ happen to be equal, the corresponding terms in (153) can be combined. In such cases, therefore, the number of constants is effectively reduced, for, as we have noticed before, the sum of two arbitrary quantities is just one arbitrary quantity. This case of equal roots, however, is the only exception [1] to the statement that (153) is the general solution of the complementary equation. Hence we have the theorem:

THEOREM V.—*If no two of the roots* $p_1, p_2, \cdots, p_s$ *of the algebraic equation*

$$F_s(p) = 0$$

are equal, the general solution of the linear differential equation

$$F_s\left(\frac{d}{dx}\right) y = 0$$

with constant coefficients and second member zero is the sum of the s exponential functions $\alpha_1 e^{p_1 x}, \alpha_2 e^{p_2 x}, \cdots, \alpha_s e^{p_s x}$, *the* α*'s being arbitrary constants.*

Suppose, now, that with these ideas before us we return to the matter of solving (147), upon the assumptions (*a*) that we know a *particular* solution of (147), and (*b*) that no two of the roots $p_1, p_2, \cdots, p_s$ of its auxiliary equation are equal. Then we can apply Theorem II and say at once that *the general solution of (147) is obtained by adding to the particular*

---

[1] In this exceptional case the solution has a different form, which will be derived in § 62.

solution the term $\sum_{j=1}^{s} \alpha_j e^{p_j x}$, *in which the* α*'s are arbitrary constants.* In particular, when the right-hand member of (147) is exponential, as in (148), a particular solution is always known in the form (150), provided the p in (148) is not itself a root of $F_s(p)$.

As a first example, consider again the equation (151). Its complementary equation is

$$\frac{d^2y}{dx^2} + 3\frac{dy}{dx} + 2y = 0.$$

Upon assuming $y = \alpha e^{px}$ as a solution of this complementary equation, it is found to be satisfactory if p satisfies the auxiliary equation

$$p^2 + 3p + 2 = 0,$$

the roots of which are -1 and -2. Hence the solution of (151) due to zero is

$$y = \alpha_1 e^{-x} + \alpha_2 e^{-2x}.$$

Finally, making use of the particular solution (152) it may be said that the general solution of (151) is

$$y = \alpha_1 e^{-x} + \alpha_2 e^{-2x} + \tfrac{1}{2} e^{-5x}.$$

As a second example, consider the equation

$$\frac{d^2y}{dx^2} + 3\frac{dy}{dx} + 2y = e^{3ix}, \tag{154}$$

which is identical with (151), except for the occurrence of an imaginary exponent. The particular solution of this equation, found by assuming $y = Ae^{3ix}$, is

$$y = -\frac{7 + 9i}{130} e^{3ix}. \tag{155}$$

The solution of the complementary equation is

$$y = \alpha_1 e^{-x} + \alpha_2 e^{-2x},$$

as before. The general solution is therefore

$$y = \alpha_1 e^{-x} + \alpha_2 e^{-2x} - \frac{7 + 9i}{130} e^{3ix}.$$

We may draw some rather remarkable deductions from this example. Returning to the particular solution (155), we may note that, since $e^{3ix} = \cos 3x + i \sin 3x$, this particular solution may be rewritten

$$y = (-\tfrac{7}{130} \cos 3x + \tfrac{9}{130} \sin 3x) + i(-\tfrac{9}{130} \cos 3x - \tfrac{7}{130} \sin 3x). \quad (156)$$

However, since (154) satisfies all the conditions laid down in Theorem III, it follows that the real and imaginary parts of (156) are due to the corresponding parts of e^{3ix}. In other words

$$y = -\tfrac{7}{130} \cos 3x + \tfrac{9}{130} \sin 3x$$

is a particular solution of

$$\frac{d^2y}{dx^2} + 3\frac{dy}{dx} + 2y = \cos 3x,$$

and

$$y = -\tfrac{9}{130} \cos 3x - \tfrac{7}{130} \sin 3x$$

is a particular solution of

$$\frac{d^2y}{dx^2} + 3\frac{dy}{dx} + 2y = \sin 3x.$$

To get the *general* solution we need only add $\alpha_1 e^{-x} + \alpha_2 e^{-2x}$ to these particular solutions, in accordance with Theorem II.

Finally, we may wisely consider some examples of the exceptional cases to which our theory, in its present state of development, does not apply.

If, for example, instead of (151) we had

$$\frac{d^2y}{dx^2} + 3\frac{dy}{dx} + 2y = 6e^{-2x}, \quad (157)$$

the solution of the complementary equation would still be

$$y = \alpha_1 e^{-x} + \alpha_2 e^{-2x};$$

and it would still be true that by adding this to any particular solution of (157) its general solution would be obtained. But if we attempt to obtain a particular solution by the method which led to (150) — that is, by assuming the solution to have the form $y = Ae^{-2x}$ and substituting it in (157) — we are led to the equation

$$A(4 - 6 + 2) = 6$$

or

$$0 \cdot A = 6,$$

from which, of course, it is impossible to obtain A. The trouble is, that (157) has no such particular solution. In fact, the desired particular solution is $y = -6xe^{-2x}$, as we can easily see by trial, and therefore the general solution is

$$y = \alpha_1 e^{-x} + \alpha_2 e^{-2x} - 6xe^{-2x}.$$

How this particular solution was derived, however, must remain unexplained for the present.

Finally, if instead of (151) we write

$$\frac{d^2y}{dx^2} + 2\frac{dy}{dx} + y = 6e^{-5x}, \tag{158}$$

a particular solution can be gotten at once in the form (150). It is

$$y = \tfrac{3}{8}e^{-5x}.$$

But as the roots of the auxiliary equation $p^2 + 2p + 1 = 0$ are -1 and -1, the attempt to use the formula (153) as a means of obtaining the general solution of the auxiliary equation leads to

$$y = \alpha_1 e^{-x} + \alpha_2 e^{-x} = (\alpha_1 + \alpha_2)e^{-x},$$

and since $\alpha_1 + \alpha_2$ is itself no more general than a single constant, this cannot be the desired general solution. Instead, the desired general solution of the complementary equation is of the form $(\alpha_1 + \alpha_2 x)e^{-x}$, as may be seen by trial; so that the general solution of (158) is

$$y = (\alpha_1 + \alpha_2 x)e^{-x} + \tfrac{3}{8}e^{-5x}.$$

§ 56. *Linear Equations with Constant Coefficients; Trigonometric Solutions*

The particular use that was made of the principle of decomposition in dealing with (154) is capable of a valuable generalization. Suppose it is desired to solve the equation [1]

$$F\left(\frac{d}{dx}\right) y = B \cos (px - \epsilon), \tag{159}$$

in which B and all the coefficients of the polynomial F are real. By the principle of decomposition the solution of (159) is the real part of the solution of

$$F\left(\frac{d}{dx}\right) y = B [\cos (px - \epsilon) + i \sin (px - \epsilon)],$$

which may also be written

$$F\left(\frac{d}{dx}\right) y = Be^{-i\epsilon} e^{ipx}.$$

As we already know how to solve this equation by the use of Theorems II, IV and V, except in certain exceptional cases, the solution of (159) can readily be obtained. The solution of

$$F\left(\frac{d}{dx}\right) y = B \sin (px - \epsilon) \tag{160}$$

will be obtained at the same time; but of course, unless there is some reason to want it, it will be a useless by-product of the process.

Although the conclusion at which we have just arrived is an immediate deduction from Theorems II, IV and V, it is of such immense importance in the study of vibrating dynamical systems to deserve a place as the sixth of our major theorems regarding linear equations.

---

[1] For the present it is not necessary to keep the order of the equation constantly in mind. Hence we may write $F\left(\frac{d}{dx}\right)$ instead of $F_s\left(\frac{d}{dx}\right)$.

THEOREM VI.—*The real part of a particular solution of a linear differential equation due to an exponential function* $Be^{ip(x-\epsilon)}$ *is a particular solution due to* $B \cos p(x - \epsilon)$; *and the imaginary part of the solution due to the exponential function, with the i cancelled, is a particular solution due to* $B \sin p(x - \epsilon)$.

Theorem VI affords quite the simplest known method of solving (159): that is, it is actually easier to solve both (159) and (160) *together* than to solve *either* of them separately. This fact may appear odd at first sight, but it has many counterparts in higher mathematics, for it is not infrequently simpler to deal with complex quantities than with real ones. In particular, many definite integrals, the values of which can only be found with difficulty by the elementary methods of the Calculus, can be evaluated quite easily by the use of certain properties of functions of complex variables.

It is essential, however, to understand clearly just what is responsible for the success of the method in this case, and to that end two things must be especially noted. The first is the peculiar property of the exponential function e^{px}: that when the operations indicated by $F\left(\frac{d}{dx}\right)$ are carried out upon it, they merely have the effect of multiplying it by $F(p)$. Because of this property the general solution of any linear equation with constant coefficients due to a function of this form can be found in the form (150) by merely solving an algebraic equation. This is the content of Theorem IV. The second is: that the cosine and sine are the real and imaginary parts of a complex exponential function; wherefore, by the principle of decomposition, the solution due to either may be inferred from that due to the exponential.

It is this intimate relation between the trigonometric and exponential functions, coupled with the ease with which solutions can be obtained for the latter, which is responsible for the success of the method just explained.

As an example, consider the simple equation

$$\frac{dy}{dx} + 2y = \sin 3x, \tag{161}$$

in which

$$F\left(\frac{d}{dx}\right) = \frac{d}{dx} + 2.$$

Since $i \sin 3x$ is the imaginary term in e^{3ix}, the solution of (161) can be obtained from the solution of

$$\left(\frac{d}{dx} + 2\right) y = e^{3ix}.$$

This latter solution is

$$\begin{aligned} y &= \frac{e^{3ix}}{F(3i)} + \alpha\, e^{-2x} \\ &= \frac{e^{3ix}}{3i + 2} + \alpha\, e^{-2x} \\ &= \frac{(2 - 3i)\, e^{3ix}}{13} + (\alpha' + i\alpha'')\, e^{-2x}. \end{aligned}$$

Hence, writing e^{3ix} in the form $\cos 3x + i \sin 3x$ and picking out the imaginary term, the solution of (161) is found to be

$$y = \frac{2 \sin 3x - 3 \cos 3x}{13} + \alpha'' e^{-2x}.$$

PROBLEMS

1. Solve (100).

2. $\dfrac{d^2y}{dx^2} + y = 0.$

3. $\dfrac{d^2y}{dx^2} + 12y = 7\dfrac{dy}{dx}.$

4. $\dfrac{d^2r}{d\phi^2} - a^2r = 0.$

5. $\dfrac{d^4y}{dx^4} - a^4y = 0.$

6. $\dfrac{d^2v}{du^2} - 6\dfrac{dv}{du} + 13v = e^{-2u}.$

7. $\dfrac{d^2y}{dt^2} + 4\dfrac{dy}{dt} - y = \sin t.$

8. $\dfrac{d^2y}{dx^2} + 3y = \sin x + \frac{1}{3}\sin 3x.$

9. $\dfrac{d^4x}{dt^4} + 16x = e^{u}.$

10. $5\dfrac{dx}{dt} + x = \sin 3t.$

11. $\dfrac{d^4x}{dt^4} - 6\dfrac{d^3x}{dt^3} + 11\dfrac{d^2x}{dt^2} - 6\dfrac{dx}{dt} = e^{-3t}.$

12. Refer to Problem 10, § 5, and explain how the methods of the present chapter can be applied to equations of the form

$$a_s x^s \frac{d^s y}{dx^s} + a_{s-1} x^{s-1} \frac{d^{s-1}y}{dx^{s-1}} + \cdots + a_1 x \frac{dy}{dx} + a_0 y = f(x).$$

13. What sort of function must $f(x)$ be, if the equation in Problem 12 is to be altogether analogous to (148)?

14. $x^4\dfrac{d^4y}{dx^4} + x^3\dfrac{d^3y}{dx^3} - 20\,x^2\dfrac{d^2y}{dx^2} + 20\,x\dfrac{dy}{dx} = 17\,x^6.$

15. $t^4\dfrac{d^4x}{dt^4} - 2\,t^3\dfrac{d^3x}{dt^3} - 20\,t^2\dfrac{d^2x}{dt^2} + 12\,t\dfrac{dx}{dt} + 16\,x = \cos(3\log t).$

§ 57. *Electrical Circuit Equations; "Transient" and "Steady State" in the Simple Series Circuit*

The rules laid down in § 17 for the relation between current and electromotive force in simple electrical circuits lead to many linear equations of just the type with which we are now dealing. The purpose of the present section is to further illustrate the principles which we have so far developed, and in particular to give concrete physical significance to the two parts of our solution — the particular solution due to $f(x)$, and the general solution due to zero — about whose separation from one another there must still be somewhat of an air of mystery.

For this purpose we choose first the very simplest illustration possible. Suppose an inductance of unit magnitude, and a resistance of magnitude 2, are connected in series, and suppose that at the instant $t = 0$ an electromotive force of the form $\sin 3t$ is impressed upon them. Then, by the rules of § 17, we find that the differential equation governing the flow of current is

$$\frac{dI}{dt} + 2I = \sin 3t.$$

This equation is identical with (161), the general solution of which has already been found to be

$$I = \frac{2 \sin 3t - 3 \cos 3t}{13} + \alpha'' e^{-2t}.$$

We still need a boundary value by means of which to determine α''. It may readily be found by the following line of reasoning: Up to the time $t = 0$ the circuit was in idleness. That is, there was no electromotive force, and therefore no current was flowing. So *just before* $t = 0$ the value of I was zero. When the electromotive force was applied, the current either rose gradually from zero, in which case the boundary condition is obvious, or else there was a sudden jump at $t = 0$. The former is the correct assumption. For a sudden jump means a finite change of current during zero time, which is equivalent in the physical sense to an infinite value for $\frac{dI}{dt}$, and therefore for $\frac{dI}{dt} + 2I$ also. This latter, however, must equal the applied electromotive force at the instant $t = 0$ and cannot be infinite. Hence it follows that there can be no sudden jump in I at $t = 0$.

We must therefore determine α'' so that I shall be zero when $t = 0$, thus arriving at the desired particular solution

$$I = \frac{2 \sin 3t - 3 \cos 3t}{13} + \frac{3}{13} e^{-2t}. \qquad (162)$$

This solution consists of two parts: the particular solution

$\frac{2}{13}\sin 3t - \frac{3}{13}\cos 3t$ due to the electromotive force $\sin 3t$ which has no arbitrary constant in it, and is therefore entirely independent of the boundary conditions; and the solution due to zero, $\alpha'' e^{-2t}$, which takes care of the boundary value and which would be of the same *form* no matter what the magnitude of the driving force might be, though of course the particular value of the constant α'' might be something else than $\frac{3}{13}$ if we changed the electromotive force.[1] In the language of electrical engineers the first of these is called the "steady-state" term and the other the "transient" term. Let us see why these names are appropriate.

Consider first the last term of (162): that is, the "transient" one. As t gets larger and larger this term becomes smaller and smaller, until finally it is no longer of practical consequence. In a very real sense, therefore, it represents a "transient current" — one which appears at $t = 0$, but soon disappears again.

The first term, on the other hand, is the sum of a sine and a cosine function, and therefore represents a sine wave with the same period as the electromotive force, and with an amplitude $1/\sqrt{13}$. Beginning at $t = 0$, it keeps on as long as the electromotive force is sustained. Hence, after the transient has died out and conditions have become "steady," this current holds the field alone. That is why it is called a "steady-state" current.[2]

Next, let us see what would occur if, after the electromotive force had been operating for a time, it were to be discontinued. To be explicit, let the time at which this takes place be

---

[1] For example, if we had an electromotive force $2\sin 3t$ instead of $\sin 3t$, α'' would be $\frac{6}{13}$.

[2] This term "steady-state" is only applied to the particular solution due to an electromotive force of simple harmonic type, or to the superposition of the solutions due to a number of such electromotive forces which happen to be acting simultaneously. That is, we might speak of the "steady-state" current due to the force

$$E = \sin t + 3\cos 2t + 6\sin\frac{1}{\pi}(t-2);$$

but not due to the force $E = t$.

$t_0 = 17\pi/6$. The circuit is still described by the same differential terms $\frac{dI}{dt} + 2I$ as before; but we now need the solution *due to zero* only, for there is no longer any electromotive force to consider. That is, the current must satisfy the complementary equation

$$\frac{dI}{dt} + 2I = 0,$$

together with a suitable boundary condition. To obtain this boundary condition we find, by substituting $t = 17\pi/6$ in (162), that the current has a magnitude $\frac{2}{13}$ at the instant when the electromotive force is removed, and by the same argument as was used above, it cannot change abruptly. Hence from this time on the current must satisfy the differential equation

$$\frac{dI}{dt} + 2I = 0,$$

together with the boundary value $I = \frac{2}{13}$ when $t = 17\pi/6$. Its equation is therefore $I = \frac{2}{13}\, e^{\frac{17}{3}\pi - 2t}$. As time passes it becomes smaller and smaller and ultimately disappears: it, too, is a transient term; the steady-state term has disappeared. The transient terms are, then, currents which might exist in the absence of any electromotive force.

In Fig. 39 are shown, both the electromotive force assumed in this illustration, the steady-state and transient terms separately, and finally the complete current.

As a second example, let us consider the general case of an inductance, resistance and capacity in series. From the rules of § 17 we obtain for this case the differential equation

$$L\frac{d^2x}{dt^2} + R\frac{dx}{dt} + \frac{x + x_0}{C} = E(t),$$

or if x is replaced by the charge on the condenser, which is $x + x_0 = q$,

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{q}{C} = E(t). \qquad (163)$$

The general solution of this equation, too, consists of a "steady-state current" due to $E(t)$, and a "transient current" determined by the boundary conditions. To determine the

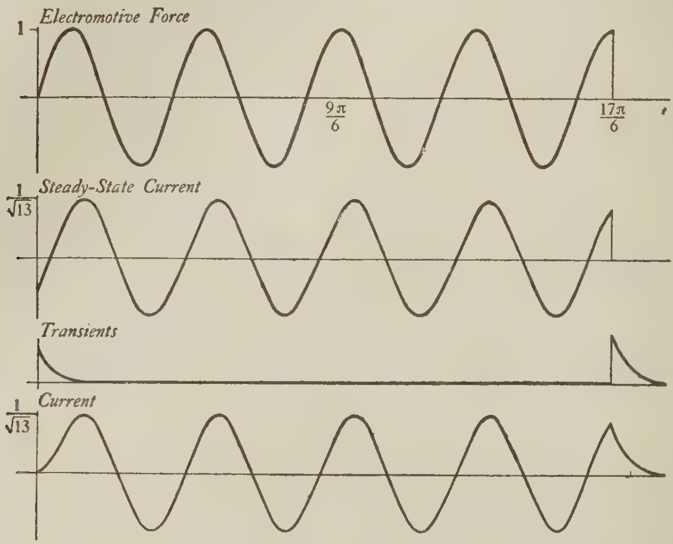

Fig. 39. — Transient and Steady-State Currents.

form of the latter, we need only find the roots of the auxiliary equation

$$Lp^2 + Rp + \frac{1}{C} = 0.$$

Obviously, they are

$$\left.\begin{aligned} p_1 &= -\frac{R}{2L} + \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}, \\ p_2 &= -\frac{R}{2L} - \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}. \end{aligned}\right\} \tag{164}$$

Then the most general solution "due to zero" — that is, the most general q which could exist in the absence of an electromotive force — is

$$q = \alpha_1 e^{p_1 t} + \alpha_2 e^{p_2 t}.$$

As I is the first derivative of x and therefore also of q, the general expression for the *transient current* is

$$I = \alpha_1 p_1 e^{p_1 t} + \alpha_2 p_2 e^{p_2 t}. \tag{165}$$

Let us see what we can learn about this transient current. Suppose, to begin with, that $R^2C < 4L$, in which case the radicals in p_1 and p_2 are imaginary. Then each term of (165) contains a factor $e^{-\frac{R}{2L}t}$ which decreases as time goes on and ultimately vanishes. The remaining factor may be denoted by $\alpha_1 p_1 e^{i\bar{n}t} + \alpha_2 p_2 e^{-i\bar{n}t}$, where $\bar{n}$ is the *real* number

$$\bar{n} = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}.$$

But as

$$e^{i\bar{n}t} = \cos \bar{n}t + i \sin \bar{n}t$$

and

$$e^{-i\bar{n}t} = \cos \bar{n}t - i \sin \bar{n}t,$$

this factor may be changed into

$$(\alpha_1 p_1 + \alpha_2 p_2) \cos \bar{n}t + i\,(\alpha_1 p_1 - \alpha_2 p_2) \sin \bar{n}t.$$

In appearance this consists of a *real* cosine term of frequency $\bar{n}/2\pi$, and an *imaginary* sine term of the same frequency. But this appearance is deceptive, for p_1 and p_2 are complex, and α_1 and α_2 may be. In fact, due to the presence of the arbitrary constants α_1, and α_2, the coefficients $\alpha_1 p_1 + \alpha_2 p_2$ and $i(\alpha_1 p_1 - \alpha_2 p_2)$ are purely arbitrary, and might as well be denoted by two new constants, a and b. In any physical problem a and b would of necessity be real, since physical quantities are never complex, but in a general mathematical sense they need not be. In either event the factor in question takes on the appearance

$$a \cos \bar{n}t + b \sin \bar{n}t.$$

Hence the entire transient solution (165) consists of a sinusoidal term of frequency $\bar{n}/2\pi$, multiplied by the amplitude factor $e^{-\frac{R}{2L}t}$. That is, the transient current is a sine wave of constantly decreasing amplitude, or as it is commonly called,

a "damped sine wave." The frequency $\bar{n}/2\pi$ with which it oscillates is known as the "natural frequency" of the circuit, for it is the frequency with which the circuit oscillates when it is allowed to oscillate "naturally," that is, without the compulsion of an imposed electromotive force.

All this is on the assumption that $R^2C < 4L$. In order to see what changes take place when this condition is violated, it is instructive to consider how the frequency $\bar{n}$ changes as R becomes larger and larger. Since

$$\bar{n} = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}},$$

it follows that $\bar{n}$ becomes smaller and smaller, and consequently the oscillation becomes slower and slower until, when $R^2C = 4L$, it ceases entirely. Another way of stating the same fact is to say that "increasing the resistance of a circuit lowers its natural frequency."

If R is increased still further the radicals in (164) become real numbers,[1] so that both p_1 and p_2 are real and negative, but not equal. Under these conditions each term of (165) is a negative exponential, which vanishes *without oscillation* as time goes on, just as did the transients of Fig. 39. This condition is generally described by saying that the circuit is "critically damped." As R is increased more and more p_2 becomes larger and larger (in absolute value), and p_1 becomes smaller and smaller. That is, increasing the resistance causes one transient component to die out more and more rapidly, and the other more and more slowly.

So far, our argument has dealt only with the transient current. But the application of Theorems IV and VI to (163) gives us also the steady-state current due to a driving force of the very important type $E = E_0 \cos nt$. Replacing $E = E_0 \cos nt$ by

$$E = E_0\, e^{int}, \tag{166}$$

[1] When R^2C is exactly equal to $4L$, p_1 and p_2 are equal, and a different form of solution is necessary, as has already been said in § 55.

of which it is the real part, (163) becomes

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{q}{C} = E_0\, e^{int},$$

a particular solution of which is

$$q = \frac{E_0\, e^{int}}{Rin + \frac{1}{C} - Ln^2},$$

or

$$I = \frac{E_0\, e^{int}}{Lin + R + \frac{1}{Cin}}. \tag{167}$$

The real part of this quantity, when added to (165), gives the most general current which may flow in the circuit in response to an alternating electromotive force of the cosine form. The separation of this real part from its attendant imaginary term is left as an exercise for the reader.

It has already been noted that the transient component dies out as time passes. Therefore, if the electromotive force is long sustained, the current in the circuit reduces sensibly to the real part of (167). This therefore is the "steady-state" current produced by the alternating driving force $E_0 \cos nt$.

§ 58. *Circuit Equations; Impedance*

It is the purpose of this section to show the intimate association between the ideas so far presented in this chapter and certain common concepts in electrical theory. For this purpose, however, it will be necessary first to outline these physical concepts.

We have seen that, in the examples considered in § 57, the current produced by a sinusoidal electromotive force was itself sinusoidal after the transient terms had disappeared. Indeed, this is true of any sort of electrical circuit which obeys a linear differential equation with constant coefficients, as we can

readily see from Theorem IV. Now, any such sinusoidal electromotive force or sinusoidal current is characterized by three properties: frequency, amplitude and phase; hence, so long as we are interested only in "steady-state" conditions we may adequately describe our circuit by telling how these three characteristics of the current are related to the corresponding characteristics of the electromotive force. All these things, we shall find, can easily be learned from a study of equations (147) to (150).

Now that we are definitely committed to an electromagnetic interpretation, however, let us replace y by I, x by t, and $f(x)$ by $E(t)$ in these equations, in order more vividly to suggest their physical meaning. Then we may say: *If any electrical circuit is of such a nature that its behavior is adequately described by the linear differential equation*

$$F\left(\frac{d}{dt}\right) I = E(t),$$

and if the electromotive force $E(t)$ is either the real or the imaginary[1] *part of an exponential function*

$$E(t) = B\, e^{int}, \tag{168}$$

then the steady-state current to which this electromotive force gives rise is the real or the imaginary part of the particular solution

$$I = \frac{B\, e^{int}}{F(in)}. \tag{169}$$

Let us first consider the equation (168) and find out what factors in the equation correspond to the terms "frequency," "amplitude" and "phase." For this purpose, however, we must separate out the imaginary part of (168); and in doing this we must give attention to the possibility that B is itself a

[1] The technical concepts with which we are about to deal were all originally developed in terms of *sine* functions. For this reason, we shall hereafter speak only in terms of *imaginary* parts of our solutions. Our argument could, of course, be carried through in terms of cosine functions if we desired; but it would require the introduction of a negative sign in one place in what would appear to be an entirely arbitrary fashion.

complex number. Suppose we write it at once in the polar form

$$B = E_0 e^{i\epsilon},$$

as we shall find it most convenient to do in the end, thus converting (168) into the form

$$E(t) = E_0 e^{i(nt+\epsilon)},$$

in which E_0, n and ϵ are all real.[1] In this form, the imaginary part is easily picked out, and is

$$E(t) = E_0 \sin (nt + \epsilon).$$

Now the reader well knows that the amplitude of such an oscillation is E_0, that it repeats itself $n/2\pi$ times per second, and that its phase is ϵ. Hence we see at once that

(*a*) If the constant B which multiplies a complex exponential such as that in (168) is written in the polar form $B = E_0 e^{i\epsilon}$, the oscillation defined by (168) has an amplitude E_0;

(*b*) It has the phase ϵ;

(*c*) Its frequency is $n/2\pi$.

Now we turn our attention to (169), and, since $F(in)$ may also be a complex number, we write it in the polar form $F(in) = Z_0 e^{i\epsilon'}$. Then we have

$$I = \frac{E_0}{Z_0} e^{i(nt+\epsilon-\epsilon')}.$$

We see at once that:

(*a*) The frequency of the current is the same as the frequency of the electromotive force.

(*b*) Its amplitude is proportional to the amplitude of the electromotive force, the factor of proportionality being the reciprocal of the absolute value of $F(in)$.

---

[1] If n were complex, its imaginary part would correspond to a damping term, as we have seen in § 57, and the electromotive force would not be "steady-state."

(*c*) Its phase is $\epsilon - \epsilon'$; that is, it differs from the phase ϵ of the electromotive force by the angle ϵ' which appears when $F(in)$ is written in the polar form.

In the study of electricity the ratio of the electromotive force to the current which it produces (that is, E/I) is called the *impedance* of the circuit. This gives us the rule that *the impedance of the circuit is* $F(in)$.

There are two other terms in common use in electrical theory for which we can easily find mathematical expressions in terms of this same function $F(in)$. To arrive at them most easily, we shall suppose that time is measured from such an instant that the phase of the current is zero. This, of course, means that $\epsilon = \epsilon'$; whence the electromotive force and current must be

$$\left.\begin{aligned} E &= E_0 \sin (nt + \epsilon'), \\ I &= \frac{E_0}{Z_0} \sin nt; \end{aligned}\right\} \qquad (170)$$

the former of which is readily thrown into the form [1]

$$E = (E_0 \cos \epsilon') \sin nt + (E_0 \sin \epsilon') \cos nt. \qquad (171)$$

In this form the electromotive force appears as the sum of two parts: one, of amplitude $E_0 \cos \epsilon'$, having the same phase as the current; the other, of amplitude $E_0 \sin \epsilon'$, differing in phase by 90°. It is customary in electrical theory to call the amplitude ratio of the in-phase component the "resistance" of the circuit, and the amplitude ratio of the out-of-phase component the "reactance" of the circuit, respectively, both these ratios being formed with the electromotive force in the numerator and the current in the denominator, as in the case

[1] If we were speaking in terms of cosine functions, the corresponding expansion would have a *negative* sign before the second term. We would then be forced to define Y (below) as the *negative* of a certain ratio. The reason is, that the out-of-phase second term in (171) *lags* 90° behind the in-phase first term; whereas, if the first term were a cosine and the second a sine, the second would *lead* the first. As the definition of reactance is set up on the basis of an electromotive force which lags behind the current produced by it, a negative sign would be necessary to correct this defect.

of the definition of impedance. Hence, denoting resistance and reactance by X and Y, respectively, we obviously have

$$X = Z_0 \cos \epsilon',$$

$$Y = Z_0 \sin \epsilon',$$

$$X + iY = Z_0 e^{i\epsilon'} = F(in).$$

In other words: *The resistance of the circuit is the real part of $F(in)$ and the reactance is its imaginary part with the i suppressed.*

The important part which the function $F(in)$ plays in this whole theory is obvious: It itself is the *impedance* of the circuit; its angle is the *phase difference* between electromotive force and current; its real and imaginary parts are the circuit's *resistance* and *reactance*, respectively. In it, in other words, are contained all four of these physical factors. Because of this, electrical engineers are coming to deal more and more in terms of the *complex quantity* $F(in)$, instead of the various real quantities into which it may be resolved. In the light of our studies we can easily define it as follows:

The impedance of a circuit which obeys the linear differential equation

$$F\left(\frac{d}{dt}\right) I = E_0 e^{int}$$

is obtained by replacing $\frac{d}{dt}$ *by* in *wherever it occurs in* $F\left(\frac{d}{dt}\right)$; *in other words, it is* $F(in)$.

PROBLEMS

1. Obtain the general solution of the circuit equation for a resistance and capacity in series. Discuss the natural frequencies.

2. Obtain the general solution of the circuit equation for an inductance only in series with a periodic electromotive force. What is the natural frequency? Discuss the types of current that might flow in the absence of any electromotive force.

3. Obtain the general solution of the circuit equation for an inductance and capacity in series with a periodic electromotive force.

What is the natural frequency? What type of current might flow in the absence of electromotive force? How long?

4. Assume, in Problem 3, that the circuit is at rest at time $t = 0$, at which the *EMF* is imposed. That is, when $t = 0$ both I and $\frac{dI}{dt}$ are zero. What sort of current flows? When is the steady state reached?

5. The *resonant* frequencies of a circuit are the frequencies for which the absolute value of the impedance $|Z|$ is a minimum. These are the real physical frequencies at which a given force will produce a greater steady-state current than for any nearby frequency. Find the resonant frequencies for the general circuit discussed in § 57. Are they the same as the natural frequencies?

6. If an open circuit containing a charged condenser, an inductance and a resistance in series, is connected to an oscillograph and then closed and an oscillogram taken, will the frequency of oscillation be the natural frequency or the resonant frequency? How would you go about obtaining the other experimentally?

7. A constant electromotive force is applied at time $t = 0$ to the general circuit of § 57. Find the current upon the assumption that the circuit was at rest until that time.

(Hint: A constant E_0 may be written $E_0 e^{0 \cdot t}$.)

8. It is not necessary that the $E(t)$ on the right-hand side of our differential equation should be capable of a simple algebraic expression. It might, for example, be represented by the pair of straight lines shown in Fig. 40. Find the current that would flow through a resistance and inductance in series, if the electromotive force were of this form.

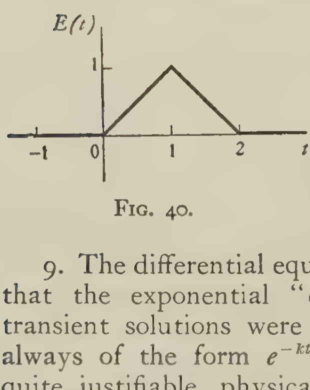

FIG. 40.

9. The differential equations used in § 57 were all of such a nature that the exponential "damping factors" which occurred in the transient solutions were all negative: that is, the solutions were always of the form $e^{-kt}\cos \bar{n}t$ rather than $e^{+kt}\cos \bar{n}t$. This was quite justifiable, physically, since stable dynamical systems never give rise to the latter type of solution. Suppose, however, that a circuit *did* obey the differential equation

$$\frac{dI}{dt} - I = E(t).$$

Discuss the current that would be produced by an electromotive force $E = \sin t$ applied at time $t = 0$.

§ 59. *Operational Methods; Factorization of Operators*

There are in mathematics, not only symbols for quantities, but symbols of operation as well. Thus 2, 3, x, n are numbers or quantities, while $+$, $-$, $\frac{d}{dx}$, $\int \cdots dx$, are obviously symbols of operation, or *operators*.[1] Mathematical methods which transform symbols of operation from one form to another are known as *operational methods*. That is, operational methods are those which involve manipulation of operators rather than manipulation of quantities.

The operator which is of most fundamental importance in the study of differential equations is $\frac{d}{dx}$, which may be denoted by p when it is desired to call special attention to its operational character.[2] As it denotes a differentiation with respect to the independent variable, two such differentiations in succession would naturally be indicated by pp, or p^2. In operational notation, therefore, a second derivative $\frac{d^2y}{dx^2}$ would be written p^2y, a third derivative p^3y, etc. Likewise any linear differential equation with constant coefficients would be written

$$F(p)\, y = f(x). \tag{172}$$

This equation means exactly the same thing as (147); the p is merely shorthand notation for $\frac{d}{dx}$. Yet it looks very much

---

[1] They may be distinguished from one another by the simple rule that symbols of quantity, like nouns, have a definite significance when standing alone; while an operator must have its meaning completed by an accompanying quantity, just as a transitive verb requires an object. Thus 2 or x represent complete ideas, just as "house" and "monkey" do; while $+$ or $\frac{d}{dx}$ do not until symbols of quantity are attached to them in some such form as $x + 2$ or $\frac{dy}{dx}$, just as "make" does not, though "make noise" does.

[2] In pure mathematics a more usual symbol is D.

like the *algebraic* equation (149) in which p was merely a number. Broadly speaking, the essence of the operational method is to treat the equation *as if p were a number;* but of course only in so far as it is possible to prove that this procedure leads to correct answers.

Now, how would (172) be solved for y if p *were* a number? Why, simply by dividing the equation through by $F(p)$ to give

$$y = \frac{f(x)}{F(p)}. \tag{173}$$

By analogy, this is called the "operational solution" of (172). The symbol $\frac{1}{F(p)}$ is called the "inverse" of the operator $F(p)$, and is often written $F^{-1}(p)$. It is a sort of generalized sign of integration and indicates the performance of all the processes necessary to solve (172).

For example, in the simplest possible case where $F(p) = p$, (172) is

$$\frac{dy}{dx} = f(x),$$

and (173) may be written either as

$$y = \frac{f(x)}{p}, \tag{174}$$

or as

$$y = \int f(x)\, dx. \tag{175}$$

Both mean *exactly the same thing*. From one point of view each is a command to integrate $f(x)$, and either command is futile unless it can be obeyed. Yet in the case of the ordinary integral notation it is altogether customary to think of the command as already having been performed and to talk of (175) as the *result* of the integration, whether it can be obtained or not. It is obviously just as permissible to so regard (174) also, for it is merely (175) written in another sort of notation. If we adopt the same attitude toward the less familiar form

(173), we come to the conclusion that, since the solution of (172) has been called $\frac{f(x)}{F(p)}$, $\frac{f(x)}{F(p)}$ is the solution of (172).[1]

Perhaps an illustration may give point to the idea. The equation

$$\frac{dy}{dx} - y = \frac{1}{x}$$

may be written as

$$(p - 1)\, y = \frac{1}{x}.$$

Similarly its solution may be expressed in either the usual form

$$y = \alpha\, e^x + e^x \int \frac{1}{x} e^{-x}\, dx, \tag{176}$$

or, in the operational form,

$$y = \frac{1}{(p-1)} \frac{1}{x}. \tag{177}$$

Neither one can be expressed in terms of the elementary functions, for $\frac{e^{-x}}{x}$ has no elementary integral; it is mere sophistry to call (176) "the solution of the differential equation" *unless* there exists some way of getting a numerical answer; and it is *not* mere sophistry to call (177) a "solution" *if* there exists a way of evaluating it. Nor are these assertions affected by the fact that one notation is familiar and the other not. Once we have a way of using it, we have just as much right to call one a solution as the other. It is exactly to the finding of such

---

[1] This argument is singularly like the one by which the Lord High Executioner in *The Mikado*, who has released a prisoner and then made affidavit to the fact of his execution, justifies his perjury:

It's like this: when your Majesty says, "Let a thing be done," it's as good as done,— practically, it is done, — because your Majesty's will is law. Your Majesty says, "Kill a gentleman," and a gentleman is told off to be killed. Consequently, that gentleman is as good as dead; practically, he is *dead, and if he is dead, why not say so?*

It is also singularly like the statement that the solution of the equation $y = 10^x$ is $x = \log_{10} y$.

Whether either is absurd depends entirely on what can be accomplished by such paralogism.

methods of interpreting and using operational solutions that the next few sections will be devoted. Of course, by ways of evaluating operational solutions we mean *right* ways; that is, ways that give correct answers; for methods which led to incorrect answers, or which *sometimes* led to incorrect answers, would be worse than no methods at all, unless we were aware of their limitations. So we shall be forced to spend a part of our time in assuring ourselves that what we are doing is not merely plausible, but right.

To begin with, we shall prove

THEOREM VII. — *When dealing with linear differential equations having constant coefficients, the operator p obeys the ordinary algebraic laws of addition, subtraction and multiplication.*

The proof that the first two laws are obeyed is almost trivial:

$$\alpha \frac{dy}{dx} + \beta \frac{dy}{dx} = (\alpha + \beta) \frac{dy}{dx}$$

becomes in operational form

$$(\alpha p + \beta p)\, y = (\alpha + \beta)\, py,$$

which proves the associative law for a first derivative, and similar statements can be made if the derivatives are of higher order.

The third point is only a little more involved. If

$$u = \alpha \frac{dy}{dx} + \beta y,$$

straight differentiation shows that

$$\gamma \frac{du}{dx} + \delta u = \gamma\alpha \frac{d^2y}{dx^2} + (\gamma\beta + \delta\alpha) \frac{dy}{dx} + \delta\beta y.$$

In operational form these become

$$u = (\alpha p + \beta)\, y$$

and

$$(\gamma p + \delta)\, u = [\gamma\alpha p^2 + (\gamma\beta + \delta\alpha)\, p + \delta\beta]\, y.$$

Substituting u from the first of these into the second gives

$$(\gamma p + \delta)(\alpha p + \beta)\, y = [\gamma\alpha p^2 + (\gamma\beta + \delta\alpha)\, p + \delta\beta]\, y\,;$$

and it will be found that when the operators in the first term are multiplied algebraically they give the second. This is the required "proof," for it says that what we get by treating the operational factors algebraically is right. Of course, it applies only to *linear* factors, and only to *two* of them; but the extensions to the product of a polynomial of any degree by a linear factor, and then to two polynomials is mere routine. In particular if s linear operators are multiplied, a polynomial operator of the sth order will result.

What is more important, however, is that this process can be reversed and an sth order expression factored in s first order terms. For suppose such an expression

$$F_s(p) = a_s p^s + a_{s-1} p^{s-1} + \cdot \cdot \cdot + a_1 p + a_0$$

has been factored into [1]

$$a_s(p - p_1)(p - p_2) \cdot \cdot \cdot (p - p_s), \qquad (178)$$

just as if it were algebraic. It is known to be permissible to recombine these linear factors by direct multiplication, and as this legitimate process produces $F_s(p)$, (178) must be equivalent to $F_s(p)$ in an operational as well as an algebraic sense. Thus we have

THEOREM VIII. — *A linear differential operator of order s with constant coefficients may be replaced by s linear differential operators of the first order, just as if it were an algebraic expression.*

For example, the operator $p^2 + 1$ can be factored into $(p + i)(p - i)$; therefore the same result should be obtained by adding to a function its own second derivative, as from first

---

[1] It must be clearly understood that the $p_1, p_2, \cdot \cdot \cdot, p_s$ are not operators. They are the numbers, real or complex, which are found as the roots of $F_s(p) = 0$. It is only the p *without a subscript* which represents $\dfrac{d}{dx}$.

forming the quantity $\frac{dy}{dx} - iy$, and then adding i times this quantity to its own derivative. The reader will find his ideas somewhat clarified by actually performing these operations upon a few functions, such as $y = x^2$, $y = \sin nx$, and the like.

The next point which should be noted in connection with this process of factorization is, that the order in which the linear factors are written is immaterial. This follows at once from the fact that (178) has been shown to be a satisfactory operational form of $F_s(p)$, whenever its factors are satisfactory in an algebraic sense, and in an algebraic sense the order of the factors is known to be immaterial. In other words, the commutative law of multiplication is established.

§ 60. *Operational Methods; The Use of Factorization in Solving Linear Equations*

To understand the use which may be made of factorization, consider the differential equation

$$\frac{d^2y}{dx^2} + y = e^x, \tag{179}$$

or

$$(p^2 + 1)\, y = e^x.$$

Assume this to be rewritten as

$$(p + i)(p - i)\, y = e^x,$$

and introduce the symbol y_1 for the quantity $(p - i)\, y$. Then (179) becomes

$$(p + i)\, y_1 = e^x,$$

or, in derivative notation,

$$\frac{dy_1}{dx} + iy_1 = e^x.$$

This is a linear equation, the solution of which is

$$y_1 = \alpha\, e^{-ix} + \frac{e^x}{1 + i}.$$

But by definition

$$(p - i)\, y = y_1,$$

or, in derivative form,

$$\frac{dy}{dx} - iy = \alpha\, e^{-ix} + \frac{e^x}{1 + i}.$$

This again is a linear equation, the solution of which is

$$y = \beta e^{ix} - \frac{\alpha}{2i} e^{-ix} + \frac{e^x}{2}.$$

Thus (179) is completely solved. The same result could have been deduced by the method of § 55.

In general, (172) may be written as

$$a_s(p - p_1)(p - p_2) \cdot \cdot \cdot (p - p_s)\, y = f(x), \qquad (180)$$

and the following definitions adopted:

$$(p - p_s)\, y = y_s,$$
$$(p - p_{s-1})\, y_s = y_{s-1},$$
$$\cdot \cdot \cdot \quad \cdot \cdot \cdot \quad \cdot \cdot \cdot$$
$$(p - p_2)\, y_3 = y_2.$$

Then (180) becomes

$$a_s(p - p_1)\, y_2 = f(x).$$

In derivative notation, each of these is a linear equation. In particular, the last is

$$\frac{dy_2}{dx} - p_1 y_2 = \frac{f(x)}{a_s}$$

and has the general solution

$$y_2 = e^{p_1 x}\left(\alpha_1 + \frac{1}{a_s}\int e^{-p_1 x} f(x)\, dx\right). \qquad (181)$$

The next preceding one is

$$\frac{dy_3}{dx} - p_2 y_3 = y_2,$$

and therefore

$$y_3 = e^{p_2 x}\left(\alpha_2 + \int e^{-p_2 x} y_2\, dx\right).$$

But as y_2 is already known from (181), this may be rewritten

$$y_3 = e^{p_2x}\left(\alpha_2 + \int e^{(p_1-p_2)x}\left(\alpha_1 + \frac{1}{a_s}\int e^{-p_1x} f(x)\,dx\right)dx\right).$$

By continuing this process y will finally be found in the form

$$y = e^{p_sx}\left(\alpha_s + \int e^{-p_sx} y_s\,dx\right)$$

$$= e^{p_sx}\left(\alpha_s + \int e^{(p_{s-1}-p_s)x}\left(\alpha_{s-1} + \cdots \left(\alpha_1 + \frac{1}{a_s}\int e^{-p_1x} f(x)\,dx\right)\cdots\right)dx\right). \tag{182}$$

Thus a new method of solving linear equations has been found. Moreover, it can be used in finding the solution due to any function $f(x)$, *not merely in finding those due to exponential functions, as was the case in* § 55. The only requirement is, that the indicated integrations can be carried out.

As an example of a solution that could not be obtained by the methods of §§ 55 and 56, consider the equation

$$\frac{d^2y}{dx^2} + 3\frac{dy}{dx} + 2y = x^2. \tag{183}$$

This can be written

$$(p+2)(p+1)\,y = x^2. \tag{184}$$

Calling $(p+1)\,y = y_2$, (184) becomes $(p+2)\,y_2 = x^2$, whence, in derivative notation, (184) is equivalent to the two first order equations

$$\frac{dy_2}{dx} + 2y_2 = x^2,$$

$$\frac{dy}{dx} + y = y_2.$$

The solution of these equations in order gives

$$y_2 = \alpha_1 e^{-2x} + e^{-2x}\int x^2 e^{2x}\,dx$$

and

$$y = \alpha_2 e^{-x} + e^{-x}\int y_2 e^{x}\, dx$$
$$= \alpha_2 e^{-x} - \alpha_1 e^{-2x} + e^{-x}\int e^{-x}\int x^2 e^{2x}\, dx\, dx.$$

This expression is easily evaluated.

It can easily be established that the s arbitrary α's in (182) are all independent. Hence *(182) may be regarded as a formula representing the general solution of any linear equation with constant coefficients.* If we separate it into two parts — one containing all the terms in which the α's appear, and the other the term

$$\frac{1}{a_s} e^{p_s x}\int e^{(p_{s-1} - p_s)x}\int \cdots \int e^{(p_1 - p_2)x}\int e^{-p_1 x} f(x)\, (dx)^s \qquad (185)$$

which has no α's — it is at once obvious that (185) must be the particular solution to which reference is made in Theorem II, and the remaining terms must constitute the general solution of the complementary equation. Now this is important; for our present results do not in any way require that all the roots p_j be distinct, as did the results stated in Theorem V. They could, therefore, be used to supplement Theorem V and find our solution "due to zero," even when repeated roots occurred. Actually, however, we shall find the algebra a bit less involved if we postpone this phase of the argument until § 62.

We must remember, then, that both (182) and (185) are true whether the p_j's are distinct or not. One gives the *general* solution of (172); the other a *particular* solution. If, however, the p_j's happen to be distinct, we can add to our *particular* solution (185) the terms required by Theorem V, and thus arrive at a form of the general solution which is somewhat simpler in appearance than (182). It is

$$y = \sum_{j=1}^{s} \alpha_j e^{p_j x} + \frac{e^{p_s x}}{a_s}\int e^{(p_{s-1} - p_s)x}\int \cdots \int e^{-p_1 x} f(x)\, (dx)^s. \qquad (186)$$

The only essential difference between (182) and (186) is,

that the latter is true *only if no two p_j's are equal*, while the former is always true.

§ 61. *Operational Methods; The Use of Partial Fractions in Solving Linear Equations*

There are other algebraic transformations of $F_s(p)$ which lead to solutions of complicated linear equations, just as factorization did in § 60. For example, $\dfrac{1}{F_s(p)}$ may be split into partial fractions. That is, the operational solution

$$y = \frac{1}{F_s(p)} f(x) \tag{173}$$

may properly be rewritten as [1]

$$y = \left(\frac{c_1}{p - p_1} + \frac{c_2}{p - p_2} + \cdots + \frac{c_s}{p - p_s}\right) f(x),$$

or, in shorthand notation,

$$y = \sum_{j=1}^{s} \frac{c_j}{p - p_j} f(x). \tag{187}$$

Of course, if such an expansion is to be valid in *general*, it must be valid in the *simplest* case: that is, when $F_s(p)$ is of the first degree. But then there is only *one* term in the partial fraction expansion, which becomes merely

$$y_j = \frac{c_j}{p - p_j} f(x). \tag{188}$$

The differential equation to which this corresponds is

$$\frac{dy_j}{dx} - p_j y_j = c_j f(x), \tag{189}$$

and its solution is

$$y_j = e^{p_j x}\left(\alpha_j + c_j \int e^{-p_j x} f(x)\, dx\right). \tag{190}$$

[1] In this section it is assumed that no two roots of $F_s(p)$ are equal. For the case of repeated roots, see § 62.

Hence if (187) is to have a correct interpretation in general, each term must be replaced by an expression of the form (190).

Suppose, then, that y is *assumed* to be the sum of s terms, each like (190), and that this sum

$$y = \sum_{j=1}^{s} y_j \tag{191}$$

is substituted in the left-hand side of (172). If (191) is a solution, the result of the substitution should be equal to $f(x)$. What is actually obtained, of course, is

$$\sum_{j=1}^{s} F_s(p)\, y_j, \tag{192}$$

and the problem before us is therefore to show that this is actually equal to $f(x)$.

With the aid of what is already known about operational methods, this is not difficult. For, in the first place, the $F_s(p)$ in each term of (192) may be factored and the factors written in any desired order (Theorem VIII). In particular, the factor $p - p_1$ may be written last (that is, next to y_1) in the first term; the factor $p - p_2$ last in the second (so that it is next to y_2), and so on. Suppose that this is done, and that the expressions

$$(p - p_1)\, y_1, \quad (p - p_2)\, y_2, \quad \cdots,$$

are replaced by the quantities

$$c_1 f(x), \quad c_2 f(x), \quad \cdots,$$

to which, by (188), they are known to be equal. Then (192) becomes

$$\begin{aligned}
&[c_1 a_s(p - p_2)(p - p_3) \cdots (p - p_s) \\
&+ c_2 a_s(p - p_1)(p - p_3) \cdots (p - p_s) \\
&\quad \cdots \qquad \cdots \qquad \cdots \qquad \cdots \\
&+ c_s a_s(p - p_1)(p - p_2) \cdots (p - p_{s-1})] f(x).
\end{aligned} \tag{193}$$

This is the result actually obtained by substituting (191) in (172). If it can be shown to equal $f(x)$, it will follow that

(191) is indeed a solution. Now it is known to be allowable to combine the terms inside the brackets just as if they were algebraic, hence it is sufficient to show that this bracketed expression reduces to unity when treated as an algebraic polynomial.

To show that it does reduce to unity, go back to (187), where $\frac{1}{F_s(p)}$ was supposed expanded in partial fractions. It was there assumed that

$$\frac{1}{F_s(p)} = \frac{c_1}{p - p_1} + \frac{c_2}{p - p_2} + \cdots + \frac{c_s}{p - p_s} \qquad (194)$$

is a true *algebraic* equation. Let it be multiplied, member by member, by

$$F_s(p) = a_s(p - p_1)(p - p_2) \cdots (p - p_s).$$

The result is obviously unity on the left side, while the right side becomes the bracketed term in (193). It follows that *the partial fraction expansion of* $\frac{1}{F_s(p)}$ *really leads to a solution of (172), provided every term of the form* $\frac{c_j}{p - p_j}f(x)$ *is replaced by the* y_j *given in (190). Moreover, this solution is the general solution of (172), for it involves s independent constants.*

It is a simple matter to write the general solution to which this process leads. It is

$$y = \sum_{j=1}^{s} \alpha_j e^{p_j x} + \sum_{j=1}^{s} c_j e^{p_j x} \int e^{-p_j x} f(x)\, dx. \qquad (195)$$

Like (186) it requires s separate integrations; but unlike (180) they are not "repeated integrals." In both cases, however, the roots p_j must first be known.

The application of this theory to particular problems is generally much simpler than its explanation has been. As an example, consider the differential equation

$$(p + 2)(p + 1)y = x^2, \qquad (184)$$

which has already been solved in § 60. Here

$$F_2(p) = (p+2)(p+1)$$

and therefore

$$\frac{1}{F_2(p)} = \frac{1}{p+1} - \frac{1}{p+2}.$$

By adding together the solutions of

$$(p+1)\,y_1 = x^2$$

and

$$(p+2)\,y_2 = -\,x^2,$$

the general solution is obtained at once in the form

$$y = \alpha_1 e^{-x} + \alpha_2 e^{-2x} + e^{-x}\int e^{x}x^2\,dx - e^{-2x}\int e^{2x}x^2\,dx. \qquad (196)$$

When the integrals are evaluated they lead to the same result as before.

As the final outcome of §§ 60 and 61 the following theorem may be stated:

Theorem IX. — *When a linear equation with constant coefficients is written in the operational form $F_s(p)\,y = f(x)$, it is permissible to subject the operator to factorization; or to subject the formal solution $y = \dfrac{f(x)}{F_s(p)}$ to partial fraction disintegration, provided that every partial fraction thus obtained is interpreted as the solution of the linear differential equation to which it itself corresponds.*

So far, of course, the theorem has only been shown to be true provided all the roots p_i are distinct, but this restriction will be removed in § 62.

PROBLEMS

1. Complete the evaluation of the two solutions of (183) found in §§ 60 and 61. Are they identical?

2. Evaluate the operational expressions:

$$(a)\quad y = \frac{x}{p - a}.$$

$$(b)\quad y = \frac{x^n}{p - a}.$$

$$(c)\quad y = \frac{\sin nx}{p - a}.$$

$$(d)\quad y = \frac{e^{nx}}{p - a}.$$

3. Solve Problems 6 and 7 of § 56 by the formula (186). Do you obtain the same solutions as before?

4. Solve the same problems by the formula (195). Which method do you prefer, the one resulting in (186) or the one resulting in (195)?

5. Solve Problem 7, § 58, by the methods of §§ 60 and 61.

6. We have seen that the transient current dies out in the circuit of Problem 5. Hence any transient necessary to satisfy boundary conditions at $t = -\infty$ must be zero at all finite times. If, however, we regard the $f(t)$ of our problem as being defined for all time by the two equations

$$f(t) = 0, \qquad t < 0,$$
$$f(t) = E_0, \qquad t > 0,$$

we will have no physical conditions to satisfy at any time except $t = -\infty$. Therefore the correct answer to our problem ought to be given by the particular solution of either (186) or (195).

Find the solution in this fashion, and see if you can check the result of Problem 5.

7. Solve Problem 8, § 58, by (186) and (195), assuming as the definition of $f(x)$ the four conditions

$$f(x) = 0, \qquad x < 0,$$
$$f(x) = x, \qquad 0 < x < 1,$$
$$f(x) = 2 - x, \qquad 1 < x < 2,$$
$$f(x) = 0, \qquad x > 2.$$

8. By parallelling the essential arguments of this chapter, develop an "Operational Theory of the Solution of the Cauchy Linear Equation."

§ 62. *Repeated Roots*

We have assumed, both in the statement of Theorem V and in the development of Theorem IX, that all the roots of the function $F_s(p)$ were distinct. Actually, Theorem V is false unless this condition is satisfied; but Theorem IX is really true in any case, as we shall now see.

To be specific, suppose three roots are equal and denote them all by p_1. Then, as is well known, the partial fraction expansion (187) should be replaced by

$$y = \frac{f(x)}{F_s(p)} = \left(\frac{c_1}{p-p_1} + \frac{c_2}{(p-p_1)^2} + \frac{c_3}{(p-p_1)^3} + \cdots\right) f(x), \quad (197)$$

in which the unwritten terms corresponding to the roots $p_4, p_5 \cdots p_s$, are of the same form as before. It is reasonable to suppose that the first three terms should represent solutions of the equations

$$\left.\begin{aligned} (p - p_1)\, y_1 &= c_1 f(x), \\ (p - p_1)^2\, y_2 &= c_2 f(x), \\ (p - p_1)^3\, y_3 &= c_3 f(x); \end{aligned}\right\} \quad (198)$$

that is, of

$$\frac{dy_1}{dx} - p_1 y_1 = c_1 f(x),$$

$$\frac{d^2y_2}{dx^2} - 2p_1 \frac{dy_2}{dx} + p_1^2 y_2 = c_2 f(x),$$

$$\frac{d^3y_3}{dx^3} - 3p_1 \frac{d^2y_3}{dx^2} + 3p_1^2 \frac{dy_3}{dx} + p_1^3 y_3 = c_3 f(x);$$

and this may actually be verified by the same argument as was used in § 61. For, upon substituting the sum of these solutions in (172), we get

$$F(p)\, y = F(p)\, y_1 + F(p)\, y_2 + F(p)\, y_3 + \cdots,$$

and (172) will be satisfied if the right-hand member of this equation sums up to $f(x)$. However, when we separate $F(p)$

into its factors, and take account of the relations (198) which the y's satisfy by definition, we get

$$\begin{aligned} F(p)\,y = [&a_s c_1 (p-p_1)^2 (p-p_4)(p-p_5)\cdots(p-p_s) \\ &+ a_s c_2 (p-p_1)(p-p_4)(p-p_5)\cdots(p-p_s) \\ &+ a_s c_3 (p-p_4)(p-p_5)\cdots(p-p_s)+\cdots]\,f(x). \end{aligned}$$

This, therefore, must be a true *differential* relation. Moreover, by Theorem VII, the differential expression to which it reduces upon combining the bracketed terms algebraically is also true. But upon multiplying (197), which is by definition a true *algebraic* equation, by $F(p)$, we find that this bracketed term reduces to unity. In other words $F\left(\frac{d}{dx}\right)y$ is indeed equal to $f(x)$. It follows, then, that Theorem IX is still true, even in the case of repeated roots.

As for the evaluation of the expressions (198), that can most easily be carried out by means of the formula (182), which we know to be valid *whether or not the roots* p_1 *are equal.* It gives us at once

$$\begin{aligned} y_1 &= \alpha_1 e^{p_1 x} + c_1 e^{p_1 x} \int e^{-p_1 x} f(x)\,dx, \\ y_2 &= (\alpha_1 x + \alpha_2) e^{p_1 x} + c_2 e^{p_1 x} \int\!\!\int e^{-p_1 x} f(x)\,(dx)^2, \\ y_3 &= (\alpha_1 x^2 + \alpha_2 x + \alpha_3) e^{p_1 x} + c_3 e^{p_1 x} \int\!\!\int\!\!\int e^{-p_1 x} f(x)\,(dx)^3. \end{aligned}$$

Upon adding together the three expressions for y_1, y_2 and y_3 we get the portion of the solution of our equation which arises from the repeated root. It is [1]

$$\begin{aligned} y = {} & (\alpha_1 x^2 + \alpha_2 x + \alpha_3)\, e^{p_1 x} \\ & + c_1 e^{p_1 x} \int e^{-p_1 x} f(x)\,dx + c_2 e^{p_1 x} \int\!\!\int e^{-p_1 x} f(x)\,(dx)^2 \\ & + c_3 e^{p_1 x} \int\!\!\int\!\!\int e^{-p_1 x} f(x)\,(dx)^3. \end{aligned}$$

---

[1] The α's, of course, represent different constants in each equation.

The portion of this result which is of primary interest is that which contains the arbitrary constants. If Theorem V had been used blindly, this portion would have been

$$(\alpha_1 + \alpha_2 + \alpha_3)\, e^{p_1 x},$$

in which there is essentially only one constant, not three. Instead, the correct result is of the form

$$(\alpha_1 x^2 + \alpha_2 x + \alpha_3)\, e^{p_1 x},$$

in which the constants are independent.

The integral terms are also modified, as compared with (195) in that the sign of integration is repeated a number of times equal to the degree of the denominator of the partial fraction to which the term corresponds.

Now this is a perfectly general result: for the solution required in the case of a fraction of the form

$$\frac{f(x)}{(p - p_1)^r}$$

may be found by simply replacing all the p_i's in (182) by p_1, and s by r. This has the effect of cancelling the exponents $p_{r-1} - p_r, p_{r-2} - p_{r-1}, \cdots, p_1 - p_2$; so that when the parentheses are removed it becomes [1]

$$y = e^{p_1 x}\left(\alpha_r + \alpha_{r-1}\int dx + \alpha_{r-2}\int\int (dx)^2 + \cdots + \alpha_1 \int\int \cdots \int (dx)^{r-1}\right)$$
$$+ e^{p_1 x}\int\int \cdots \int e^{-p_1 x} f(x)\,(dx)^r.$$

The last term of this is indeed of the same form as the integral terms of (195), except that an r-fold integration is demanded. As for the other term, it integrates at once into

$$e^{p_1 x}\left(\alpha_r + \alpha_{r-1} x + \frac{\alpha_{r-2}}{2!} x^2 + \cdots + \frac{\alpha_1}{(r-1)!} x^{r-1}\right)$$

[1] Note that a_r — that is, the coefficient of p^r in the polynomial $(p - p_1)^r$ — is unity.

which we might as well write

$$e^{p_1x}(\alpha_r + \alpha_{r-1}x + \alpha_{r-2}x^2 + \cdots + \alpha_1 x^{r-1}),$$

since the α's are quite arbitrary. Hence we have the theorems:

THEOREM X. — *A particular solution of the differential equation*

$$(p - p_1)^r y = f(x)$$

is

$$e^{p_1x}\int\int\cdots\int e^{-p_1x}f(x)(dx)^r.$$

THEOREM XI. — *If a root p_1 of the equation $F_s(p) = 0$ in Theorem V is repeated r times, the general solution of the corresponding differential equation*

$$F_s\left(\frac{d}{dx}\right)y = 0$$

contains the exponential e^{p_1x} multiplied, not by one arbitrary constant, but by a polynomial $(\alpha_1 x^{r-1} + \alpha_2 x^{r-2} + \cdots + \alpha_r)$ in which there are r such constants.

As an example, consider the equation

$$(p + 1)(p + 1)(p + i)\,y = x \qquad (199)$$

in which the root -1 occurs twice. Upon splitting $\frac{1}{F_3(p)}$ into partial fractions we obtain

$$\frac{1}{(p+1)^2(p+i)} = \frac{(1+i)^2}{4(p+i)} - \frac{(1+i)^2}{4(p+1)} - \frac{1+i}{2(p+1)^2}.$$

Hence, if the expressions $\frac{x}{p+1}$, $\frac{x}{(p+1)^2}$ and $\frac{x}{p+i}$ can be interpreted, the solution of (199) can be found. But $\frac{x}{p+1}$ must be the solution of

$$\frac{dy}{dx} + y = x,$$

which is

$$\alpha e^{-x} + e^{-x}\int e^{x} x\, dx = \alpha e^{-x} + x - 1.$$

Similarly $\dfrac{x}{p+i} = \gamma e^{-ix} + 1 - ix.$ On the other hand, $\dfrac{x}{(p+1)^2}$ is the solution of

$$(p+1)(p+1)\, y = x,$$

which works out to be

$$(\alpha + \beta x)\, e^{-x} + e^{-x}\int\!\!\int e^{x} x\, (dx)^2 = (\alpha + \beta x)\, e^{-x} + x - 2.$$

Adding all these results together, the particular solution due to x is found to be

$$\frac{(1+i)^2}{4}(1 - ix) - \frac{(1+i)^2}{4}(x-1) - \frac{1+i}{2}(x-2) = 1 + 2i - ix, \quad (200)$$

while the general solution due to zero is

$$(\alpha + \beta x)\, e^{-x} + \gamma e^{-ix}. \quad (201)$$

The sum of (200) and (201) is the general solution of the equation.

In dealing with this example we have actually carried out the algebraic transformations leading to (201). This, however, was not necessary; for we knew in advance as a result of our general argument that it must take the form which we finally obtained.

§ 63. *The Exceptional Case of the Exponential Solution*

We have still one exceptional situation to consider, for we have seen that Theorem IV is only true provided $F_s(p) \neq 0$. But though Theorem IV is not true in this case, the methods of §§ 60 to 62 can still be applied, and equation (182) as well as Theorems IX and X are still true. By their aid we may deal with this difficulty also.

As an example, consider the equation

$$\frac{dy}{dx} + y = e^{-x},$$

or, in operational form, $(p + 1)\,y = e^{-x}$. By either (78), (182) or (195) the solution of this equation is

$$y = \alpha e^{-x} + e^{-x}\int e^{x}\cdot e^{-x}\,dx$$

$$= \alpha e^{-x} + xe^{-x}.$$

Before commenting on this expression, consider the further example

$$\frac{d^3y}{dx^3} + 3\,\frac{d^2y}{dx^2} + 3\,\frac{dy}{dx} + y = e^{-x},$$

or, in symbolic form, $(p + 1)^3\,y = e^{-x}$. By Theorems X and XI, its solution is

$$y = (\alpha_1 x^2 + \alpha_2 x + \alpha_3)e^{-x} + e^{-x}\int\int\int (dx)^3$$

$$= (\alpha_1 x^2 + \alpha_2 x + \alpha_3)e^{-x} + \frac{x^3}{3!}\,e^{-x}.$$

These solutions differ in form from those which would be expected if p were not a root of the auxiliary equation only in the inclusion of a power of x in the particular solution. The change is therefore similar to that which occurs in the complementary part of the solution when roots are repeated. But the important thing to note is, that the operational *methods* built up in §§ 60, 61 and 62 apply even when the exponential method of § 55 fails.

PROBLEMS

1. $\dfrac{d^3y}{dx^3} - \dfrac{d^2y}{dx^2} - \dfrac{dy}{dx} + y = 0.$

2. $(p^4 - 3p^3 + 3p^2 - p)\,y = e^{2x}.$

3. $(p^3 - p^2 + p - 1)\,r = \cos\theta.$

4. Discuss the types of transient current which may flow in the circuit of equation (163) when $R^2C = 4L$. Draw a curve representing a typical case.

5. Solve equation (90), § 33.

6. What form of function, in the case of the Cauchy equation is analogous to the exceptional case of the exponential solution discussed in § 63?

7. Extend your essay of Problem 8, § 61, to cover the case of repeated roots in the Cauchy equation.

8. Solve the equation

$$x^2 \frac{d^2y}{dx^2} + 3x \frac{dy}{dx} + y = \frac{1}{x}.$$

9. Solve the differential equation (199) by the use of formula (182).

§ 64. *A Peculiar Property of (186) and (195)*

In § 21, and again in § 25, we called attention to the fact that the arbitrary constant which arises in connection with an indefinite integral, such as (51) or (56), can equally well be written as a limit of integration as in (52) and (63). This fact does not depend at all upon the nature of the differential equation which gave rise to the integral, but is perfectly general; for example, the constant of integration which appears in our formula (78) for the solution of a linear differential equation of the first order might just as well be written as the lower limit of integration, thus throwing the formula into the form

$$y = e^{-\int f_1 dx} \int_{x_0}^{x} e^{\int f_1 dx} f_2 \, dx,$$

in which x_0 plays the same rôle of arbitrary constant as was played by the α of equation (78).[1] Similarly, we can add

[1] We could add limits of integration to the integral signs which occur in the exponents if we wished to do so. But while this would do no harm, it would also do no good, because upon referring to the derivation of (78) it will be found that the same arbitrary constant that was added to *one* of the exponents would have to be subtracted from the *other*, and thus the two would cancel.

limits of integration to (182) to take the place of the solution of the complementary equation. When we do this (182) takes the form

$$y = \frac{e^{p_s x}}{a_s} \int_{x_1}^{x} e^{(p_{s-1} - p_s)x} \int_{x_2}^{x} \cdots \int_{x_s}^{x} e^{-p_1 x} f(x)(dx)^s, \quad (202)$$

in which the arbitrary constants are $x_1, x_2, \cdots, x_s$. This is the general solution of the differential equation

$$F_s(p)\, y = f(x),$$

just as truly as is (182), and from it any particular solution can be obtained by assigning an appropriate set of values to the arbitrary lower limits of the integrals.

Now it is a remarkable property of this form of expression that

Theorem XII. — *If all the lower limits of integration in (202) are assigned the same numerical value (which we may call x_0), the solution satisfies the following set of boundary values: that not only y but its first $s - 1$ derivatives as well shall all vanish at the point $x = x_0$.*

That this is true follows at once upon writing down the various derivatives of (202), and then replacing x by x_0. The equations, however, are rather complicated and we shall not attempt to reproduce them.

Instead we turn our attention to (195), which represents the general solution as it was obtained from the process of partial fraction expansion. We can again say that the arbitrary constants of integration can be replaced by arbitrary lower limits of the integrals, thus leading to the formula

$$y = \sum_{j=1}^{s} c_j e^{p_j x} \int_{x_j}^{x} e^{-p_j x} f(x)\, dx. \quad (203)$$

Moreover, it is again true that

Theorem XIII. — *If the lower limits of integration in*

(203) are all equal to x_0, y and its first $s - 1$ derivatives vanish for this value of x.

We shall carry out the proof of this theorem. To begin with, however, we need certain relationships between the constants c_j and p_j which appear in (195), and these we can derive most easily from (194).

Since $F_s(p)$ is known to be of the sth degree, the Taylor's series for $\frac{1}{F_s(p)}$ must be of the form

$$\frac{1}{a_s p^s} + \frac{k}{p^{s+1}} + \frac{l}{p^{s+2}} + \cdots. \tag{204}$$

That is, all the terms with exponents smaller than s must disappear. On the other hand, if we expand the terms on the right-hand side of (194) in descending powers of p and add together the various series which we get in this way, the sum must be this same series (204) ; for it is a well-known theorem that the Taylor's series for any function is unique. Suppose we carry out this process. From the first term we obtain

$$\frac{c_1}{p - p_1} = \frac{c_1}{p} + \frac{c_1 p_1}{p^2} + \frac{c_1 p_1^2}{p^3} + \cdots,$$

from the second

$$\frac{c_2}{p - p_2} = \frac{c_2}{p} + \frac{c_2 p_2}{p^2} + \frac{c_2 p_2^2}{p^3} + \cdots,$$

and so on. From the fact that the sum of all of them must be identical with (204) we conclude at once that the coefficients of all the terms for which the exponent of p is less than s must sum up to zero. This gives us the desired relationships

$$\left.\begin{aligned} c_1 + c_2 + \cdots + c_s &= 0, \\ c_1 p_1 + c_2 p_2 + \cdots + c_s p_s &= 0, \\ c_1 p_1^2 + c_2 p_2^2 + \cdots + c_s p_s^2 &= 0, \\ \cdots \quad \cdots \quad & \cdots \quad \cdots \\ c_1 p_1^{s-2} + c_2 p_2^{s-2} + \cdots + c_s p_s^{s-2} &= 0. \end{aligned}\right\} \tag{205}$$

We are now ready to prove our theorem: that is, that the y defined by

$$y = \Sigma y_j = \sum_{j=1}^{s} c_j e^{p_j x} \int_{x_0}^{x} e^{-p_j x} f(x)\, dx \qquad (206)$$

vanishes at $x = x_0$ together with its first $s - 1$ derivatives.

That y itself vanishes is immediately obvious, for when $x = x_0$ the limits of integration become equal and therefore every term in the sum vanishes. To see that the derivatives vanish we make use of (189) and (191). From these it follows that

$$\frac{dy}{dx} = \Sigma p_j y_j + f(x)\, \Sigma c_j,$$

or remembering the first of equations (205),

$$\frac{dy}{dx} = \Sigma p_j y_j. \qquad (207)$$

We have already called attention, however, to the fact that every term y_j in (206) vanishes at $x = x_0$. Therefore, (207) must also vanish at $x = x_0$.

Proceeding one step further we differentiate (207) and obtain

$$\frac{d^2y}{dx^2} = \Sigma p_j \frac{dy_j}{dx} = \Sigma p_j^2 y_j + f(x)\, \Sigma c_j p_j,$$

the last term of which again drops out because of the second of of equations (205), while the first term vanishes at $x = x_0$ because every y_j is zero there.

It is obvious that we could continue this process step by step as long as the relations (205) hold out: for the third derivative we would have to make use of the third of relations (205), for the fourth derivative we would have to make use of the fourth equation, and so on. Therefore we conclude that as many derivatives must vanish at $x = x_0$ as there are equations in the set (205), and this number is obviously $s - 1$; hence Theorem XIII is proved.

As an illustration, consider again the equation (183) and its solution (196). According to Theorem XIII, we can cause

both this solution and its first derivative to vanish at $x = 0$ by dropping entirely the terms which contain arbitrary constants and adding the limits 0 and x to the integrals. When this is done and the integrations are actually carried out they lead to

$$y = \frac{x^2}{2} - \frac{3x}{2} + \frac{7}{4} - 2e^{-x} + \frac{1}{4}e^{-2x}.$$

The reader can readily assure himself of the fact that both this expression and its first derivative really do vanish at $x = 0$.

§ 65. *Non-Integrable Functions*

The general idea of treating differential operators as if they were algebraic quantities can often be used to attain valuable series expansions of integrals which cannot be evaluated in finite form. We may illustrate the general line of attack by the consideration of one or two special examples:

In § 59 attention was called to the fact that neither (176) nor its operational form (177) could be integrated in finite form. On the other hand, in Problems 1 and 2, § 25, the reader has already found that (176) could be expanded in either of two series, one in the ascending powers of x and the other in descending powers. Similar results may also be obtained directly from (177).

If p were an algebraic quantity, $\frac{1}{p-1}$ could be expanded in either of the two forms

$$\frac{1}{p-1} = -1 - p - p^2 - p^3 - \cdots,$$

or

$$\frac{1}{p-1} = \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \cdots.$$

If it is assumed that the same process is legitimate in the case of the operator, it follows that, in the first case

$$y = -\frac{1}{x} - p\frac{1}{x} - p^2\frac{1}{x} - \cdots,$$

and in the second

$$y = \frac{1}{p}\frac{1}{x} + \frac{1}{p^2}\frac{1}{x} + \frac{1}{p^3}\frac{1}{x} + \cdots.$$

However, the terms of the first series are the successive derivatives of $\frac{1}{x}$, so that, in series form,

$$y = -\frac{1}{x} + \frac{1}{x^2} - \frac{1\cdot 2}{x^3} + \frac{1\cdot 2\cdot 3}{x^4} - \cdots,$$

which agrees with the answer to Problem 2, § 25, except that the arbitrary term αe^x has been omitted.

Similarly, the terms of the second series are

$$\frac{1}{p}\frac{1}{x} = \int \frac{1}{x}\,dx = \log x,$$

$$\frac{1}{p^2}\frac{1}{x} = \int \log x\,dx = x(\log x - 1),$$

$$\frac{1}{p^3}\frac{1}{x} = \int x(\log x - 1)\,dx = \frac{x^2}{2!}\left(\log x - 1 - \frac{1}{2}\right),$$

and in general

$$\frac{1}{p^{n+1}}\frac{1}{x} = \frac{x^n}{n!}\left(\log x - 1 - \frac{1}{2} - \frac{1}{3} - \cdots - \frac{1}{n}\right).$$

By adding all these together, and noting that the coefficient of $\log x$ is identical with the series for e^x, the result is found to be

$$y = e^x \log x - x - \frac{x^2}{2!}\left(1 + \frac{1}{2}\right) - \frac{x^3}{3!}\left(1 + \frac{1}{2} + \frac{1}{3}\right) - \cdots.$$

This result is different from that obtained in Problem 1, § 25, though both involve ascending powers of x.

It is entirely possible to justify the use of operational expansions such as these in a wide variety of problems, but the attempt to do so would be entirely beyond the limits of this textbook. We must, however, warn the reader that the method has its limitations and sometimes gives incorrect results when applied to situations for which it is not suited. He

should, therefore, never use it without either assuring himself in advance of the fact that it must of necessity give him a correct result, or else definitely establishing the correctness of his answer after it has been obtained.

PROBLEMS

1. Solve Problem 4, § 58, by the method of § 64.

2. Solve Problem 2, § 63, subject to the boundary conditions that both y and its first three derivatives shall vanish at $x = 1$.

3. The differential equation of Problem 8, § 58, has an operational solution of the form $\dfrac{f(x)}{Lp + R}$. It ought, therefore, to be possible to treat it by expanding this operator in a series in either ascending or descending powers as in § 64. Upon attempting to do so, however, the reader will find that while one of these methods leads to a correct solution, the result obtained from the other one is incorrect.

Plot the various approximation curves, as well as the true solution which they should approach.

CHAPTER VIII

Systems of Linear Equations

§ 66. *General Outline of the Method of Solving Systems of Linear Equations*

The operational methods of §§ 60–65 can easily be extended so as to apply to *systems* of linear equations. To do this, we need only recall that the essential idea back of the operational scheme is that of regarding the symbol $\frac{d}{dx}$ as if it were an algebraic quantity, carrying out our solution on this assumption, and then seeking some method of interpreting the operational answer thus obtained.

It will probably make the general discussion somewhat simpler if we first illustrate the processes with which we are to deal by means of an example. With this in mind, consider the pair of equations

$$\left.\begin{aligned} \frac{d^2y_1}{dx^2} + 3y_1 + \frac{dy_2}{dx} - \quad y_2 &= x, \\ -\frac{dy_1}{dx} - 9y_1 + \frac{dy_2}{dx} + 5y_2 &= 0. \end{aligned}\right\} \tag{208}$$

If we replace the differential symbol $\frac{d}{dx}$ by p and treat it as if it were algebraic, we arrive at the equivalent operational equations

$$(p^2 + 3)\, y_1 + (p - 1)\, y_2 = x,$$

$$-(p + 9)\, y_1 + (p + 5)\, y_2 = 0.$$

If these equations were algebraic and we desired to solve them

for the variables y_1 and y_2, we would probably use the method of determinants and write the solutions in the form

$$\left.\begin{aligned} y_1 &= \frac{M_1}{\Delta}x, \\ y_2 &= \frac{M_2}{\Delta}x, \end{aligned}\right\} \tag{209}$$

where M_1, M_2 and Δ stand for the three determinants

$$\left.\begin{aligned} M_1 &= \begin{vmatrix} 1 & p-1 \\ 0 & p+5 \end{vmatrix} = (p+5), \\ M_2 &= \begin{vmatrix} p^2+3 & 1 \\ -p-9 & 0 \end{vmatrix} = (p+9), \\ \Delta &= \begin{vmatrix} p^2+3 & p-1 \\ -p-9 & p+5 \end{vmatrix} = (p+1)(p+2)(p+3). \end{aligned}\right\} \tag{210}$$

Our next step must be to seek some way of interpreting these results. Suppose we try the expedient, which proved to be satisfactory in connection with the single linear equation, of expanding the operational solution in terms of partial fractions. We readily find that

$$\left.\begin{aligned} \frac{p+5}{(p+1)(p+2)(p+3)} &= \frac{2}{p+1} + \frac{-3}{p+2} + \frac{1}{p+3}, \\ \frac{p+9}{(p+1)(p+2)(p+3)} &= \frac{4}{p+1} + \frac{-7}{p+2} + \frac{3}{p+3}; \end{aligned}\right\} \tag{211}$$

whence our operational solutions (209) become

$$\left.\begin{aligned} y_1 &= \frac{2}{p+1}x - \frac{3}{p+2}x + \frac{1}{p+3}x, \\ y_2 &= \frac{4}{p+1}x - \frac{7}{p+2}x + \frac{3}{p+3}x. \end{aligned}\right\} \tag{212}$$

We learned in § 61 that when we were dealing with a single equation it was legitimate to interpret any operational expres-

sion, such as (188), as the solution of the corresponding linear differential equation (189). This led us to the general result (190). It is natural to try placing the same interpretation upon the six terms of (212) But we are now confronted by a question which did not arise in the case of a single equation. Since both y_1 and y_2 in (209) have the same denominator Δ, the two equations (212) are actually composed of the same three terms $\frac{1}{p+1}x$, $\frac{1}{p+2}x$ and $\frac{1}{p+3}x$, only the numerical coefficients being different. We believe that each of these should be interpreted as the solution of the simple equation to which it corresponds. The question which arises, however, is: Shall we or shall we not require $\frac{1}{p+1}x$ to be *the same* solution in both y_1 and y_2? If we do, we shall have three arbitrary constants — one for each of the three terms — but *the same three* will appear in both equations. If we do not, the arbitrary constants in the second equation will be independent of those in the first. In one case we shall have *three* arbitrary constants; in the other *six*.

Obviously, our wisest course is to assume the second assumption to be the true one, for it leads to the more general result. If it proves wrong, we may then investigate the other.

We easily find that

$$\frac{1}{p+1}x = \alpha_1 e^{-x} + (x - 1),$$

$$\frac{1}{p+2}x = \alpha_2 e^{-2x} + \frac{1}{2}\left(x - \frac{1}{2}\right),$$

$$\frac{1}{p+3}x = \alpha_3 e^{-3x} + \frac{1}{3}\left(x - \frac{1}{3}\right),$$

the α's being the arbitrary constants. If we follow our second assumption — that the constants in the various terms are entirely unrelated — we arrive at the alleged solutions

$$y_1 = 2\alpha_1 e^{-x} - 3\alpha_2 e^{-2x} + \alpha_3 e^{-3x} + \tfrac{5}{6}x - \tfrac{49}{36},$$

$$y_2 = 4\alpha_4 e^{-x} - 7\alpha_5 e^{-2x} + 3\alpha_6 e^{-3x} + \tfrac{3}{2}x - \tfrac{31}{12}.$$

We must now determine whether or not these *alleged* solutions are *correct.* This can most easily be done by direct substitution in the original differential equations (208). When we do this, however, we find that the left-hand side of the first equation reduces to

$$8(\alpha_1 - \alpha_4)e^{-x} - 21(\alpha_2 - \alpha_5)e^{-2x} + 12(\alpha_3 - \alpha_6)e^{-3x} + x \quad (213)$$

(not x), while the left-hand side of the second equation reduces to

$$-16(\alpha_1 - \alpha_4)e^{-x} + 21(\alpha_2 - \alpha_5)e^{-2x} - 6(\alpha_3 - \alpha_6)e^{-3x} \quad (214)$$

(not zero). In other words, *our operational process does not lead us to a true solution of the equations* if we follow our second hypothesis and allow a different constant of integration to be introduced in every term. On the other hand, if $\alpha_1 = \alpha_4$, $\alpha_2 = \alpha_5$, $\alpha_3 = \alpha_6$, (213) and (214) do reduce to x and 0, as they ought. The true solution is therefore

$$y_1 = 2\alpha_1 e^{-x} - 3\alpha_2 e^{-2x} + \alpha_3 e^{-3x} + \tfrac{5}{6}x - \tfrac{49}{36},$$
$$y_2 = 4\alpha_1 e^{-x} - 7\alpha_2 e^{-2x} + 3\alpha_3 e^{-3x} + \tfrac{3}{2}x - \tfrac{31}{12}.$$

That is, *our operational solution gives a correct result if each partial fraction is given the same meaning, arbitrary constant and all, wherever it occurs.*

So far, of course, this statement is known to be true only of the particular example with which we are dealing, and our next step must be to assure ourselves of the fact that it is true in general. We may simplify our argument somewhat, however, by digressing long enough to introduce an artifice by means of which any system of equations can be reduced to another of the first order, and a generalization of our principle of superposition.

§ 67. *Another Principle of Superposition*

Let us consider two systems of equations: one the system

$$\left.\begin{array}{l} F_{11}(p)\,y_1 + F_{12}(p)\,y_2 + \cdots + F_{1r}(p)\,y_r = f_1(x), \\ F_{21}(p)\,y_1 + F_{22}(p)\,y_2 + \cdots + F_{2r}(p)\,y_r = f_2(x), \\ \cdots \qquad \cdots + \cdots \qquad \cdots \qquad \cdots \\ F_{r1}(p)\,y_1 + F_{r2}(p)\,y_2 + \cdots + F_{rr}(p)\,y_r = f_r(x)\,; \end{array}\right\} \quad (215)$$

in which the F's are any linear operators whatever, and the other a system identical with (215) except that $f_1(x), f_2(x), \cdots, f_r(x)$ are replaced by another set of functions $g_1(x), g_2(x), \cdots, g_r(x)$. That is, we suppose we are dealing with any two identical systems, except that in one case we demand solutions "due to" certain functions f_i, while in the other we demand solutions due to certain other functions g_i. Let us suppose, moreover, that we know a set of solutions for each of them :

$$y_1 = \phi_1(x),\ y_2 = \phi_2(x),\ \cdots,\ y_r = \phi_r(x)$$

being the set due to f_i and

$$y_1 = \psi_1(x),\ y_2 = \psi_2(x),\ \cdots,\ y_r = \psi_r(x)$$

that due to g_i.

Then we may easily prove the following theorem :

THEOREM XIV. — *The solutions of a system of equations due to the set of functions $f_i + g_i$ are the sums of the solutions due to f_i and those due to g_i separately.*

To prove this, we need only notice that if we substitute the quantities

$$\begin{aligned} y_1 &= \phi_1 + \psi_1, \\ y_2 &= \phi_2 + \psi_2, \\ &\cdots \\ y_r &= \phi_r + \psi_r \end{aligned}$$

in the left-hand side of any one of equations (215) — say the first — we can rearrange the terms in the order

$$[F_{11}(p)\phi_1 + F_{12}(p)\phi_2 + \cdots + F_{1r}(p)\phi_r] \\ + [F_{11}(p)\psi_1 + F_{12}(p)\psi_2 + \cdots + F_{1r}(p)\psi_r].$$

However, the first bracket is equal to $f_1(x)$ by definition, and the second to $g_1(x)$. Similarly, had we substituted the same y's in the second of equations (215) we would have gotten $f_2 + g_2$, and so on. This proves our theorem.

Now if Theorem XIV is true of *two* sets of functions f_i and g_i, it is true of any number. In particular, it would be true of the sets

$$\begin{array}{ccccc} f_1, & 0, & 0, & \cdots, & 0; \\ 0, & f_2, & 0, & \cdots, & 0; \\ 0, & 0, & f_3, & \cdots, & 0; \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & 0, & \cdots, & f_r; \end{array}$$

from which we conclude

THEOREM XV. — *The system (215) can be solved by adding together the solutions of the following r sets of equations:* (1) *a set in which every f except* f_1 *is replaced by zero;* (2) *a set in which every f except* f_2 *is replaced by zero;* $\cdots$; (*r*) *a set in which every f except* f_r *is replaced by zero.*

This theorem enables us to simplify our general study somewhat: for whatever we can prove to be true of a system in which every f but one is zero can be applied at once to the more general case.

§ 68. *The Relation between a Single Linear Equation and a System of Equations*

Suppose that in the differential equation (183) we arbitrarily introduce the notation $\dfrac{dy}{dx} = y_1$. Then (183) becomes identical with the system of two equations

$$\left.\begin{aligned} \frac{dy_1}{dx} + 3y_1 + 2y &= x^2, \\ y_1 - \frac{dy}{dx} &= 0, \end{aligned}\right\} \qquad (216)$$

and it requires no great amount of effort to prove that when this set of equations is treated by the method used in § 66 it

leads to the same solution as that obtained in § 60. Similarly, by introducing the notation $y_1 = \frac{dy}{dx}, y_2 = \frac{d^2y}{dx^2}, \cdots$, the general linear equation (146) can be reduced to the system of s equations

$$\left.\begin{array}{r} a_s \dfrac{dy_{s-1}}{dx} + a_{s-1} y_{s-1} + \cdots + a_1 y_1 + a_0 y = f(x), \\ y_{s-1} - \dfrac{dy_{s-2}}{dx} \qquad\qquad\qquad = 0, \\ \cdots \qquad \cdots \qquad \cdots \qquad \cdots \\ y_1 - \dfrac{dy}{dx} = 0, \end{array}\right\} \qquad (217)$$

and it would be found, on carrying out the process of the last section, that both this system and the original equation (146) possessed exactly the same solution.

Now the noteworthy thing about both (216) and (217) is, that no equation in either of them contains a derivative of higher order than the first. In other words, by the expedient of calling the derivatives by other names we have succeeded in replacing a linear equation of any order whatever by a system of linear equations, each of which is of the first order.

The same device can be used to reduce a *system* of equations to the first order. For example, the introduction of the symbol y_3 for $\frac{dy_1}{dx}$ reduces the set of two equations (208), one of which is of the second order, to the equivalent set of three first order equations

$$\begin{aligned} \frac{dy_3}{dx} + \frac{dy_2}{dx} - y_2 + 3y_1 &= x, \\ y_3 \qquad\qquad\qquad - \frac{dy_1}{dx} &= 0, \\ -y_3 + \frac{dy_2}{dx} + 5y_2 - 9y_1 &= 0. \end{aligned}$$

It follows, then, that any linear equation or any system of

linear equations, no matter what their order, can always be reduced to a system of the first order.

From the standpoint of the practical solution of differential equations, there is nothing to be gained by this transformation. We can usually solve problems in one form just as easily as in the other; and indeed in those cases where there is anything to be said in favor of the one method rather than the other, it is usually in favor of not making the transformation. From a theoretical standpoint, however, the device is a very useful one, for because of it we may conclude that whatever is true of sets of equations of the first order must also be true of those of higher order from which they arose.

§ 69. *Proof of the Correctness of the General Operational Solution*

We have established the fact that, provided we are able to solve the set of equations [1]

$$\left.\begin{aligned}
&(a_{11}p + a'_{11})\,y_1 + (a_{12}p + a'_{12})\,y_2 \\
&\qquad\qquad + \cdots + (a_{1r}p + a'_{1r})\,y_r = f(x), \\
&(a_{21}p + a'_{21})\,y_1 + (a_{22}p + a'_{22})\,y_2 \\
&\qquad\qquad + \cdots + (a_{2r}p + a'_{2r})\,y_r = 0, \\
&(a_{r1}p + a'_{r1})\,y_1 + (a_{r2}p + a'_{r2})\,y_2 \\
&\qquad\qquad + \cdots + (a_{rr}p + a'_{rr})\,y_r = 0,
\end{aligned}\right\} \tag{218}$$

which are of the first order and in which only one f appears, we can solve the general set (215). We have reason to believe also, as a result of the example considered in § 66, that the solution can be found by the method of partial fractions provided we give an operator of the form $\dfrac{1}{p - p_i} f(x)$ exactly the same meaning wherever it occurs. We shall see that this is really true.

[1] It need hardly be said that some of the a's and a''s may be zero.

Let us solve (218) algebraically by the method of determinants. We thus get

$$\left.\begin{aligned} y_1 &= \frac{M_1}{\Delta} f, \\ y_2 &= \frac{M_2}{\Delta} f, \\ \cdots & \quad \cdots \\ y_r &= \frac{M_r}{\Delta} f, \end{aligned}\right\} \tag{219}$$

in which Δ is the determinant of the system (218) and M_1, $M_2, \cdots, M_r$ are the minors of the elements in its first row.

We now assume for the purposes of our discussion that Δ is of degree s in p,[1] and that no two of its roots $p_1, p_2, \cdots, p_s$ are equal.[2] Finally, we assume that $M_1, M_2, \cdots, M_r$ are all of lower degree [3] than Δ.

Under these assumptions, any quotient M_k/Δ may be expanded in the form

$$\frac{M_k}{\Delta} = \frac{c_{1k}}{p - p_1} + \frac{c_{2k}}{p - p_2} + \cdots + \frac{c_{sk}}{p - p_s}, \tag{220}$$

in which the essential thing to remember is, that the operators $\dfrac{1}{p - p_1}, \dfrac{1}{p - p_2}, \cdots, \dfrac{1}{p - p_s}$ are the same in every such expansion, though the c's by which they are multiplied may be different for the different M's.

---

[1] Unless the a's and a''s happen to be such that the highest powers of p cancel, Δ will be of degree r. In any event $s \leqq r$.

[2] The purpose of this assumption is merely to simplify our discussion. The modifications introduced by repeated roots are exactly similar to those met in Chapter VII.

[3] They obviously are if Δ is of degree r. In case $s < r$, however, it is possible for one or more M's to be of equal or greater degree than Δ. If so, the partial fraction expansion of M/Δ will begin with a constant or with a positive power of p. We could modify our proof to include these cases also; but it is probably best to have it understandable even though some generality is sacrificed in making it so.

The reader will find some degenerate equations of this sort among the problems at the end of the chapter. Indeed, the Δ of Problem 6 does not even contain p.

We now agree to substitute these expansions in (219), thus getting

$$y_1 = c_{11}\frac{f}{p-p_1} + c_{21}\frac{f}{p-p_2} + \cdots + c_{s1}\frac{f}{p-p_s},$$

$$y_2 = c_{12}\frac{f}{p-p_1} + c_{22}\frac{f}{p-p_2} + \cdots + c_{s2}\frac{f}{p-p_s},$$

$$\cdots \qquad \cdots \qquad \cdots \qquad \cdots$$

$$y_r = c_{1r}\frac{f}{p-p_1} + c_{2r}\frac{f}{p-p_2} + \cdots + c_{sr}\frac{f}{p-p_s};$$

and to give to $\dfrac{f}{p-p_i}$ *exactly the same interpretation in each of these equations.* That is, we shall write as our *alleged* solutions

$$\left.\begin{aligned} y_1 &= \sum_{j=1}^{s} c_{j1}\psi_j(x), \\ y_2 &= \sum_{j=1}^{s} c_{j2}\psi_j(x), \\ &\cdots \qquad \cdots \qquad \cdots \\ y_r &= \sum_{j=1}^{s} c_{jr}\psi_j(x), \end{aligned}\right\} \tag{221}$$

where ψ_j is found by solving

$$\frac{d\psi_j}{dx} - p_j\psi_j = f(x) \tag{222}$$

to be

$$\psi_j = \alpha_j e^{p_j x} + e^{p_j x}\int_0^x e^{-p_j x} f(x)\, dx. \tag{223}$$

We shall prove that these alleged solutions really satisfy (218) by actually substituting them in these equations, and showing that they are satisfied. For this purpose, however, we shall need values for $\dfrac{dy_1}{dx}, \cdots, \dfrac{dy_r}{dx}$. From (221) we see that all these derivatives have the same form

$$\frac{dy_k}{dx} = \sum_{j=1}^{s} c_{jk}\frac{d\psi_j}{dx};$$

or, if we replace $\dfrac{d\psi_j}{dx}$ by the value $p_j\psi_j + f(x)$ deduced from (222),

$$\frac{dy_k}{dx} = \sum_{j=1}^{s} c_{jk}p_j\psi_j + f(x)\sum_{j=1}^{s} c_{jk}. \tag{224}$$

Finally, from (221) and (224) we get

$$a\frac{dy_k}{dx} + a'y_k = \sum_{j=1}^{s}(ap_j + a')c_{jk}\psi_j + f(x)\sum_{j=1}^{s} ac_{jk}.$$

We now go back to the first of equations (218) and note that every term on the left-hand side is of just this form. Hence, when our alleged solutions are substituted in this first equation they give

$$\left.\begin{aligned} L_1 &= \sum_{k=1}^{r}\left(a_{1k}\frac{dy_k}{dx} + a'_{1k}y_k\right) \\ &= \sum_{k=1}^{r}\sum_{j=1}^{s}(a_{1k}p_j + a'_{1k})\,c_{jk}\psi_j + f(x)\sum_{k=1}^{r}\sum_{j=1}^{s} a_{1k}c_{jk}. \end{aligned}\right\} \tag{225}$$

Similarly, when we substitute our alleged solutions in the left-hand member of the second of equations (218) we get

$$\left.\begin{aligned} L_2 &= \sum_{k=1}^{r}\left(a_{2k}\frac{dy_k}{dx} + a'_{2k}y_k\right) \\ &= \sum_{k=1}^{r}\sum_{j=1}^{s}(a_{2k}p_j + a'_{2k})\,c_{jk}\psi_j + f(x)\sum_{k=1}^{r}\sum_{j=1}^{s} a_{2k}c_{jk}. \end{aligned}\right\} \tag{226}$$

The remaining equations could be thrown into similar forms.

These quantities $L_1, L_2, \cdots$, are the values which the left-hand members actually take when the alleged solutions are substituted in them. If the alleged solutions are true solutions, L_1 must be equal to $f(x)$, while L_2 and the other L's must all be zero. The problem before us is to show that the L's really have these values.

To do this, we turn to our determinant

$$\Delta = \begin{vmatrix} a_{11}p + a'_{11}, & a_{12}p + a'_{12}, & \cdots, & a_{1r}p + a'_{1r} \\ a_{21}p + a'_{21}, & a_{22}p + a'_{22}, & \cdots, & a_{2r}p + a'_{2r} \\ \cdots & \cdots & \cdots & \cdots \\ a_{r1}p + a'_{r1}, & a_{r2}p + a'_{r2}, & \cdots, & a_{rr}p + a'_{rr} \end{vmatrix}$$

and recall that the M's are the minors of the elements in the first row. Hence, if we expand Δ in terms of this row, we get

$$\Delta = (a_{11}p + a'_{11})\, M_1 + (a_{12}p + a'_{12})\, M_2 + \cdots,$$

or

$$\sum_{k=1}^{r} \frac{M_k}{\Delta} (a_{1k}p + a'_{1k}) = 1. \qquad (227)$$

It is also known to be a general property of determinants that if these minors $M_1, M_2, \cdots, M_r$, are multiplied by the elements of any other row than the first, the sum of the products thus obtained is zero. Hence we get, from the elements of the second row,

$$\sum_{k=1}^{r} \frac{M_k}{\Delta} (a_{2k}p + a'_{2k}) = 0. \qquad (228)$$

Similar equations could be written for all the other rows.

We now replace M_k/Δ in both (227) and (228) by its value (220). The results are

$$\frac{1}{p - p_1} \sum_{k=1}^{r} c_{1k}(a_{1k}p + a'_{1k}) + \frac{1}{p - p_2} \sum_{k=1}^{r} c_{2k}(a_{1k}p + a'_{1k})$$
$$+ \cdots + \frac{1}{p - p_s} \sum_{k=1}^{r} c_{sk}(a_{1k}p + a'_{1k}) = 1 \qquad (229)$$

and

$$\frac{1}{p - p_1} \sum_{k=1}^{r} c_{1k}(a_{2k}p + a'_{2k}) + \frac{1}{p - p_2} \sum_{k=1}^{r} c_{2k}(a_{2k}p + a'_{2k})$$
$$+ \cdots + \frac{1}{p - p_s} \sum_{k=1}^{r} c_{sk}(a_{2k}p + a'_{2k}) = 0. \qquad (230)$$

Of course, there is an equation similar to (230) for every other row of Δ.

Now all these equations (220) to (230) are true for *every* value of p. Suppose, then, that we first multiply (229) by $p - p_1$, and afterward let p approach p_1. It is obvious that every term except the first must vanish, for they all contain the factor $p - p_1$. Hence in the limit, when $p = p_1$, we have

$$\sum_{k=1}^{r} c_{1k}(a_{1k}p_1 + a'_{1k}) = 0.$$

Next, let us multiply (229) by $p - p_2$ and then let p approach p_2. We get

$$\sum_{k=1}^{r} c_{2k}(a_{1k}p_2 + a'_{1k}) = 0.$$

In general, if we multiply (229) by $p - p_j$ and let p approach p_j we get

$$\sum_{k=1}^{r} c_{jk}(a_{1k}p_j + a'_{1k}) = 0.$$

Moreover, by treating (230) in the same way, we can conclude that

$$\sum_{k=1}^{r} c_{jk}(a_{2k}p_j + a'_{2k}) = 0.$$

A glance at the first terms in the right-hand members of (225) and (226) now shows, however, that the summations with respect to k are of exactly this form. Hence these terms drop out and leave us

$$L_1 = f(x) \sum_{k=1}^{r} \sum_{j=1}^{s} a_{1k} c_{jk} \tag{231}$$

and

$$L_2 = f(x) \sum_{k=1}^{r} \sum_{j=1}^{s} a_{2k} c_{jk}. \tag{232}$$

Of course, the other L's take similar forms.

Next, we turn again to (229) and let p approach infinity.

Since every term is of the form $c\,\dfrac{ap + a'}{p - p_i}$, they all reduce in the limit to terms of the form ac. Thus (229) becomes

$$\sum_{k=1}^{r} a_{1k} c_{1k} + \sum_{k=1}^{r} a_{1k} c_{2k} + \cdots + \sum_{k=1}^{r} a_{1k} c_{sk} = 1,$$

or

$$\sum_{j=1}^{s} \sum_{k=1}^{r} a_{1k} c_{jk} = 1. \tag{233}$$

By treating (230) in exactly the same way, we get

$$\sum_{j=1}^{s} \sum_{k=1}^{r} a_{2k} c_{jk} = 0, \tag{234}$$

and other equations of this latter type would be obtained from the other rows of Δ.

We now observe that the summations which still remain on the right-hand side of (231) and (232) are just (233) and (234). Hence

$$L_1 = f(x),$$

$$L_2 = 0.$$

In other words, our alleged solutions have indeed made the left- and right-hand members of (218) equal.

We may therefore state the following theorem:

Theorem XVI. — *It is legitimate to expand in partial fractions the operational solution of a system of linear differential equations with constant coefficients, provided any partial fraction* $\dfrac{1}{p - p_i} f(x)$ *is given exactly the same interpretation — constant of integration included — wherever it appears.*

Of course, our proof of this theorem is only valid so long as the roots of Δ are distinct and the minors M are of lower degree than Δ itself. The theorem itself, however, is true even when these conditions are not satisfied.

§ 70. *The Solution of a System of Differential Equations Subject to the Boundary Condition that All Variables Shall Vanish at the Same Point*

There is one set of values of the arbitrary constants in (223) which is of peculiar interest: namely, the set in which every α_j is zero. When this set of values is used every ψ vanishes at $x = 0$. Hence by (221), $y_1, y_2, \cdots, y_r$ are also zero at this point.

This property is entirely analogous to the one developed in § 64 for the case of a single equation of order higher than the first. In fact the result of § 64 follows immediately from the present argument, for, as we saw in § 68, we can reduce a single equation of high order to a set of first order equations by the expedient of introducing new variables to represent the successive derivatives. Having done this, (221) would tell us that not only y itself but all of its derivatives of order less than s would vanish at $x = 0$.

The argument of the present section, however, applies equally well to a set of linear equations of any order whatsoever, since, as we have already seen, any such set can be reduced to a new set of first order equations. To do it we need only introduce new letters for all the derivatives of every variable *except those of the highest order*. To say, then, that every y in the new system vanishes at $x = 0$ is equivalent to saying that every variable in the original system also vanished at $x = 0$, together with all of its derivatives *up to but not including* those of the highest order.[1]

§ 71. *The Special Case of $f(x) = e^{vx}$*

There are two special cases of (221) which have come to have special names and are frequently referred to in engineering

[1] This result is always true under the hypotheses with which we worked in § 69. It is even true if the roots of Δ are not all distinct. But unlike Theorem XV it is not always valid when the other hypothesis is violated; that is, when one or more M's are of as high degree as Δ. In such cases the partial fraction expansions may have certain terms which require no integration and which do not vanish when $x = 0$.

Among the problems at the end of § 74 are some examples of this sort.

literature. The present section will consider the more general one of the two, which is known as the *Generalized Heaviside Expansion Theorem*, while in the next section we shall consider the more special case that is known as the *Heaviside Expansion Theorem.*

If we assume the function $f(x)$ on the right-hand side of (218) to be an exponential Be^{px} and require our solution to satisfy the special set of boundary conditions discussed in the § 70, (223) integrates at once into

$$\psi_i = B\frac{e^{px} - e^{p_j x}}{p - p_i}.$$

Suppose now that we substitute this in (221), and then break it up into two sums, the one containing all those terms in which the exponential e^{px} occurs, the other containing the remaining terms. The result is

$$y_k = Be^{px}\sum_{j=1}^{s}\frac{c_{jk}}{p - p_i} - B\sum_{j=1}^{s}\frac{c_{jk}\,e^{p_j x}}{p - p_i}. \tag{235}$$

Now let us go back to (220) and notice that the summation which occurs in the first term of (235) is exactly $M_k(p)/\Delta(p)$, p meaning here of course the number which occurs in the exponential e^{px}. Introducing this in (235), we have the final formula

$$y_k = B\frac{M_k(p)}{\Delta(p)}e^{px} - B\sum_{j=1}^{s}\frac{c_{jk}e^{p_j x}}{p - p_i}. \tag{236}$$

This result is exceedingly simple. Not only is the term M_k/Δ obtained by merely replacing the *differential symbol* p in (218) by the *number* p, and then solving the system of equations as if it were algebraic instead of differential; but even the second term of (236) is obtained by expanding M_k/Δ in a series of partial fractions, and then multiplying each one of these partial fractions by the exponential $e^{p_j x}$.

In other words, when the function f on the right-hand side of (218) is an exponential and when all of the variables in the desired solution are required to vanish at $x = 0$, the solution

can be obtained by purely algebraic means *without any use of the Calculus whatever.*

As an example, suppose that the right-hand side of (208) had been $\cos x$ instead of x. As this is the real part of e^{ix}, we may obtain our solution by first finding the solution due to this exponential and then keeping only the real part. For the first terms of y_1 and y_2 we set $p = i$ in (210), since i is the coefficient of x in our exponential. This gives us the terms

$$\frac{M_1}{\Delta}e^{ix} = \left(\frac{1}{10} - \frac{i}{2}\right)e^{ix},$$

$$\frac{M_2}{\Delta}e^{ix} = \left(\frac{1}{10} - \frac{9i}{10}\right)e^{ix}.$$

For the second terms of (236) we refer to equation (211), which we modify, both by introducing the particular value $p = i$ with which we are dealing, and also by multiplying each term by the exponential term to which it corresponds. This gives us

$$\frac{2e^{-x}}{i+1} + \frac{-3e^{-2x}}{i+2} + \frac{e^{-3x}}{i+3},$$

$$\frac{4e^{-x}}{i+1} + \frac{-7e^{-2x}}{i+2} + \frac{3e^{-3x}}{i+3}.$$

When these two sets of results are collected together and their real parts are sorted out, they give us the desired solutions

$$y_1 = \tfrac{1}{10}\cos x + \tfrac{1}{2}\sin x - e^{-x} + \tfrac{6}{5}e^{-2x} - \tfrac{3}{10}e^{-3x},$$

$$y_2 = \tfrac{1}{10}\cos x + \tfrac{9}{10}\sin x - 2e^{-x} + \tfrac{14}{5}e^{-2x} - \tfrac{9}{10}e^{-3x}.$$

The practical importance of (236) lies in the fact that electrical problems very often lead to the boundary conditions discussed in § 70. If, then, the electromotive force happens to be of the form $\cos pt$ or $\sin pt$, which it often is, we can obtain the particular solution which satisfies the boundary conditions by merely carrying out a certain amount of elementary algebraic manipulation. No integrations are necessary.

§ 72. *The Heaviside Theorem*

The result which is ordinarily spoken of as the Heaviside Theorem in books on electricity is a theorem entirely analogous to the one stated in § 71, but which assumes that the function $f(x)$ is a *constant* E instead of an exponential Be^{px}. That is, it corresponds to the electrical case of a *steady* electromotive force applied at a certain instant, instead of to the case of an *alternating* electromotive force.

The constant E, however, may be written $Ee^{0 \cdot x}$ as we have several times done in the past. As there is nothing about the argument of § 71 which requires that p shall be different from zero, it follows at once that the solution of our present problem can be obtained by merely writing zero in place of p wherever it occurs in (236). The result is

$$y_k = E\frac{M_k(0)}{\Delta(0)} + E\sum_{j=1}^{s}\frac{c_{jk}e^{p_jx}}{p_j}. \tag{237}$$

For example, if we were to solve (208) with the x on the right-hand side of the first equation replaced by 1, and subject to the boundary conditions that not only y_1 and y_2, but the first derivative of y_1 as well, should vanish at $x = 0$, we would obtain the first terms of (237) by replacing p by zero in M_1/Δ and M_2/Δ. This would give us $\frac{5}{6}$ and $\frac{3}{2}$, respectively. For the second terms of (237) we would set $p = 0$ in (211) and multiply each of the terms by the corresponding value of e^{p_jx}. The final result would then be

$$y_1 = \tfrac{5}{6} - 2e^{-x} + \tfrac{3}{2}e^{-2x} - \tfrac{1}{3}e^{-3x},$$

$$y_2 = \tfrac{3}{2} - 4e^{-x} + \tfrac{7}{2}e^{-2x} - e^{-3x}.$$

§ 73. *Transient and Steady State in Electrical Networks*

The applications of the laws laid down in § 17 to a circuit of more than one mesh leads to a system of linear differential equations. One example of this sort was given in § 17. Another will be presented here.

The differential equations controlling the flow of current in the network shown in Fig. 41, when expressed in terms of "quantity," are

$$L\frac{d^2x_1}{dt^2} + R\frac{dx_1}{dt} + \frac{1}{C}(x_1 - x_2) = E(t),$$

$$\frac{1}{C}(x_2 - x_1) + R'\frac{dx_2}{dt} = 0;$$

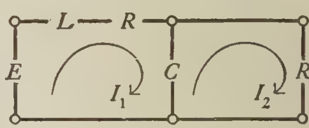

FIG. 41

or, in operational form,

$$\left(Lp^2 + Rp + \frac{1}{C}\right)x_1 - \frac{x_2}{C} = E,$$

$$-\frac{x_1}{C} + \left(R'p + \frac{1}{C}\right)x_2 = 0.$$

Their "operational solution" is therefore

$$x_1 = \frac{\left(R'p + \frac{1}{C}\right)E}{\Delta(p)},$$

$$x_2 = \frac{E}{C\Delta(p)},$$

where

$$\Delta(p) = \begin{vmatrix} Lp^2 + Rp + \frac{1}{C} & -\frac{1}{C} \\ -\frac{1}{C} & R'p + \frac{1}{C} \end{vmatrix}.$$

For the currents the solutions are

$$I_1 = \frac{dx_1}{dt} = \frac{p\left(R'p + \frac{1}{C}\right)E}{\Delta(p)},$$

$$I_2 = \frac{dx_2}{dt} = \frac{pE}{C\,\Delta(p)}.$$

Now let us assume that this circuit is in idleness until the time $t = 0$, and that then the electromotive force $E(t) = E_0 \cos nt$ is applied. By an argument entirely similar to that used in § 57 we find that at the instant $t = 0$ both x_1 and x_2, and also $\frac{dx_1}{dt}$, must be zero. These are exactly the boundary conditions laid down in § 70; and if we replace $\cos nt$ by e^{int} we may even make use of the special results of § 71.

For this purpose, however, we must know the roots of $\Delta(p)$. We easily find that

$$\Delta(p) = LR'p^3 + \left(RR' + \frac{L}{C}\right)p^2 + \frac{R' + R}{C}p,$$

whence its roots are

$$\left.\begin{aligned} p_1 &= 0, \\ p_2 &= \frac{-RCR' - L + \sqrt{(RCR' + L)^2 - 4CLR'(R' + R)}}{2CLR'}, \\ p_3 &= \frac{-RCR' - L - \sqrt{(RCR' + L)^2 - 4CLR'(R' + R)}}{2CLR'}. \end{aligned}\right\} \qquad (238)$$

We are now ready to make use of the general formula (236), which gives us

$$\left.\begin{aligned} x_1 &= E_0 \frac{R'in + \frac{1}{C}}{\Delta(in)} e^{int} - E_0 \sum_{j=1}^{3} \frac{c_{j1}}{in - p_j} e^{p_j t}, \\ x_2 &= E_0 \frac{1}{C\Delta(in)} e^{int} - E_0 \sum_{j=1}^{3} \frac{c_{j2}}{in - p_j} e^{p_j t}; \end{aligned}\right\} \qquad (239)$$

where the coefficients c_{j1} and c_{j2} could be found by expanding $\frac{R'p + \frac{1}{C}}{\Delta(p)}$ and $\frac{1}{C\Delta(p)}$ in partial fractions. They are rather complicated expressions to write, and as their exact values are of no interest to us, we can best allow them to remain undetermined.

Finally, by differentiating (239) we find the currents to be [1]

$$\left.\begin{aligned} I_1 &= E_0 \frac{in\left(R'in+\dfrac{1}{C}\right)}{\Delta(in)} e^{int} - E_0 \frac{c_{21}\,p_2}{in-p_2} e^{p_2t} - E_0 \frac{c_{31}\,p_3}{in-p_3} e^{p_3t}, \\ I_2 &= E_0 \frac{in}{C\Delta(in)} e^{int} - E_0 \frac{c_{22}\,p_2}{in-p_2} e^{p_2t} - E_0 \frac{c_{32}\,p_3}{in-p_3} e^{p_3t}. \end{aligned}\right\} \quad (240)$$

Now let us return to (238) and notice that, like the equivalent expressions (164), p_2 and p_3 may be either real and *negative* or else complex with a *negative* real part. In either case as time passes the terms e^{p_2t} and e^{p_3t} in (240) become smaller and smaller, and ultimately disappear. These terms are therefore "transients." On the other hand, the term e^{int} does not disappear as time goes on. It is the steady state to which the current eventually settles down. In other words *the first term of the generalized Heaviside expansion gives the steady-state solution of the problem; the remaining terms are transients.*

With this interpretation before us, we can write down at once a number of statements paralleling those in §§ 57 and 58, but applying now to a network of any number of meshes instead of a simple series circuit:

(*a*) The roots of $\Delta(p) = 0$ give us the *natural frequencies* of the network and also the exponential damping factors by which they are multiplied.

(*b*) No matter which one of the currents we deal with, it has the same natural frequencies as any other.

(*c*) The ratios E/I_1 and E/I_2 are called *impedances* as before. The former, which gives the current in the mesh to which the electromotive force is applied, is called an *input impedance;* the other a *transfer impedance.*

(*d*) The real parts of these complex impedances are called the *input resistance* and the *transfer resistance* of the circuit; while the imaginary parts are called *reactances.*

---

[1] The terms in p_1 disappear since the derivative of $e^{p_1t} = e^0 = 1$ is zero.

§ 74. *The Solution of Steady-State Problems*

There is one outstanding difficulty with all this theory regarding linear equations: it requires time after time that we find the roots $p_1, p_2, \cdots, p_s$ of an algebraic equation. Now this is perfectly satisfactory so long as the equation happens to be easy to solve, but when we attempt to deal with the common run of physical problems we find that many of them are far from simple. In fact, so unusual are the easy kind that a great deal of care must be taken to find problems and examples enough to illustrate our text.

There is one type of problem, however, which does not suffer from this difficulty. It is the sort in which we are interested only in that particular solution which we have called "steady-state." We have seen that such problems require only the writing down of the determinant Δ and the particular minor M in which we happen to be interested; for, by (236) and our subsequent interpretation of it in § 73,

$$E_0 \frac{M(in)}{\Delta(in)} e^{int}$$

is the steady-state solution due to the force $E_0 e^{int}$. As this process does not require us to solve the equation $\Delta(p) = 0$, it is often possible to find steady-state solutions even when the transient terms are beyond our reach.

For example, consider the network shown in Fig. 42. Its differential equations are

$$\begin{aligned} (Lp + R)\, I_1 - RI_2 &= E, \\ -RI_1 + (Lp + 2R)\, I_2 - RI_3 &= 0, \\ -RI_2 + (Lp + 2R)\, I_3 - RI_4 &= 0, \\ -RI_3 + (Lp + 2R)\, I_4 &= 0. \end{aligned}$$

Suppose, now, that we wish to know what sort of current will *eventually* (that is, after transients have disappeared) flow in the

last mesh of this circuit in response to an electromotive force $E_0 \cos nt$. We write:

$$\Delta(p) = \begin{vmatrix} Lp + R & -R & 0 & 0 \\ -R & Lp + 2R & -R & 0 \\ 0 & -R & Lp + 2R & -R \\ 0 & 0 & -R & Lp + 2R \end{vmatrix}$$

$$= L^4 p^4 + 7L^3 R p^3 + 15L^2R^2 p^2 + 10LR^3 p + R^4,$$

$$M_4(p) = -\begin{vmatrix} -R & Lp + 2R & -R \\ 0 & -R & Lp + 2R \\ 0 & 0 & -R \end{vmatrix} = R^3.$$

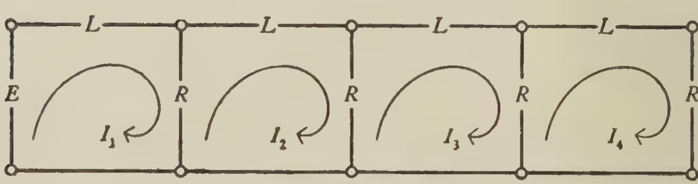

FIG. 42.

Thus we have I_4 equal to the real part of

$$\frac{E_0R^3 e^{int}}{(L^4 n^4 - 15L^2R^2 n^2 + R^4) - i\,(7L^3R n^3 - 10LR^3 n)}$$

or

$$I_4 = \frac{E_0R^3(L^4n^4 - 15L^2R^2n^2 + R^4)}{(L^4n^4 - 15L^2R^2n^2 + R^4)^2 + (7L^3Rn^3 - 10LR^3n)^2} \cos nt$$

$$- \frac{E_0R^3(7L^3Rn^3 - 10LR^3n)}{(L^4n^4 - 15L^2R^2n^2 + R^4)^2 + (7L^3Rn^3 - 10LR^3n)^2} \sin nt.$$

This much we can easily do: to find the transient solution of our problem, however, would be very difficult, as it would require the solution of a cubic equation.

PROBLEMS

1. Solve the system of equations:

$$(a)\quad \frac{d^2y}{dx^2} + y + 4z = 1,$$

$$\frac{dy}{dx} + 3\frac{dz}{dx} - z = 0.$$

$$(b)\quad \frac{d^2I_1}{dt^2} + I_1 - \frac{dI_2}{dt} - I_2 = \cos nt,$$

$$2\frac{dI_1}{dt} + \frac{dI_2}{dt} + I_2 = 0.$$

$$(c)\quad \frac{dy}{dx} + 2y - 3z = x,$$

$$y - \frac{d^2z}{dx^2} = 0.$$

$$(d)\quad (p+3)\,y_1 - (p+2)\,y_2 = f_1(x),$$
$$(p+1)\,y_1 + (p+10)\,y_2 = f_2(x).$$

2. Solve

$$(p^2-4)\,y_1 + (p^2+1)\,y_2 = f_1(x),$$
$$(p^3+6)\,y_1 + (p^3+5p)\,y_2 = f_2(x).$$

(Since Δ is of lower degree than M_1 and M_2, it is necessary to divide out the fractions M_1/Δ and M_2/Δ until a remainder of lower degree than Δ is obtained. This can then be separated into partial fractions.)

3. Solve

$$(p+1)\,y_1 + (p+3)\,y_2 = f_1(x),$$
$$p\,y_1 + (p+2)\,y_2 = f_2(x).$$

(By noting that $y_1 + y_2 = f_1 - f_2$, and substituting this in the differential equation, it is possible to obtain the solutions at once. However, it is worth while to try to carry out the formal operational process.)

CHAPTER IX

Other Equations of Order Higher than the First

§ 75. *Introductory*

We have now reached the end of what may properly be regarded as the legitimate field of an elementary text in Differential Equations; for equations of high order which are not linear, or which if linear have variable coefficients, usually require a special technique which is more properly reserved for an advanced text. Nevertheless, there are certain concepts which recur so frequently in technical literature that the reader will probably find an introduction to them of value. The purpose of this chapter is to provide such an *introduction*, not a working knowledge.

§ 76. *A Common Trick*

In § 33 certain second order equations were considered from which either x or y was absent, and it was found that they reduced immediately to first order equations when $y' = \dfrac{dy}{dx}$ was regarded as the dependent variable. There is a closely allied trick, relating to certain equations from which both x and $\dfrac{dy}{dx}$ are absent, with which the reader should be familiar because of its almost universal use in mechanics and physics. Such an equation, when solved for $\dfrac{d^2y}{dx^2}$, becomes

$$\frac{d^2y}{dx^2} = f(y). \tag{241}$$

Suppose this is multiplied by $2\frac{dy}{dx}$, thus giving

$$2\frac{dy}{dx}\frac{d^2y}{dx^2} = 2f(y)\frac{dy}{dx}. \tag{242}$$

It is obvious, from inspection, that the first member is the derivative of $\left(\frac{dy}{dx}\right)^2$, while the second member is the x-derivative of $2\int f(y)\,dy$. Hence

$$\left(\frac{dy}{dx}\right)^2 = 2\int f(y)\,dy + \alpha \tag{243}$$

and

$$x + \beta = \int\frac{dy}{\sqrt{2\int f(y)\,dy + \alpha}}$$

is the solution of (241).

When properly viewed, this trick is identical with the processes of § 33, as may be seen by carrying out the solution as would have been done in that section:

Writing $\frac{d^2y}{dx^2} = y'\frac{dy'}{dy}$, (241) becomes

$$y'\,dy' = f(y)\,dy.$$

This differs very little from (242). In fact, if it is multiplied by 2 and divided by dx, it becomes identical with (242). Its solution is obviously

$$y'^2 = 2\int f(y)\,dy + \alpha,$$

which is identical with (243); and from this point on the two methods are quite the same.

Neither method of attack is greatly to be preferred over the other: such advantage as there is lies with the method of § 33, which is capable of solving equations from which $\frac{dy}{dx}$ is not absent, whereas the other method is not. The example given in § 33 is a case at point. However, the point of view of the

present section is so widespread that the reader is certain sooner or later to come in contact with it.

§ 77. *Solution in Series*

In § 25 attention was called to the fact that differential equations frequently lead to integrals which cannot easily be evaluated, and it was shown that these integrals could often be expanded in series of one sort or another. Again in § 65 we found that the use of operational methods could lead to series solutions. The use of series, however, is really much broader than might be inferred from these examples, for they can be called into play when the solution of the equation cannot be written in the form of an explicit integral, or even an "operational solution." The second illustration of this section is of that type; while the first is merely an illustration of the method of attack.

Suppose it is desired to solve the equation

$$\frac{dy}{dx} = y^2. \tag{244}$$

The solution is obviously

$$y = -\frac{1}{x + \alpha}, \tag{245}$$

but for the time being it will be supposed to be unknown. Suppose, moreover, that the solution is *assumed* to be written in the form of a series

$$y = a_0 + a_1 x + a_2 x^2 + \cdots, \tag{246}$$

in which the a's are all unknown. *If* there is such a series and *if* it is convergent, the series for $\frac{dy}{dx}$ can easily be found by differentiating it term by term. Moreover, the series for y^2 can be found by multiplying (246) by itself. The two results are

$$\frac{dy}{dx} = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \cdots, \tag{247}$$

$$\left.\begin{aligned} y^2 = a_0^2 + 2a_0a_1 x &+ 2a_0a_2 x^2 + 2a_0a_3 x^3 + \cdots \\ &+ a_1^2 x^2 \quad + 2a_1a_2 x^3 + \cdots \end{aligned}\right\} \tag{248}$$

It can be shown that two series

$$a_0 + a_1 x + a_2 x^2 + \cdots$$

and

$$b_0 + b_1 x + b_2 x^2 + \cdots$$

never represent the same function unless every a is equal to the corresponding b: that is, unless the series are identical. Hence if (246) is really a solution of (244), (247) and (248) must be identical, for both of the latter must be equal to $\frac{dy}{dx}$. This, however, requires that

$$\begin{aligned} a_1 &= a_0^2, \\ 2a_2 &= 2a_0 a_1, \\ 3a_3 &= 2a_0 a_2 + a_1^2, \\ 4a_4 &= 2a_0 a_3 + 2a_1 a_2, \\ \cdots & \quad \cdots \quad \cdots . \end{aligned}$$

From these equations it is easily seen that

$$\begin{aligned} a_1 &= a_0^2, \\ a_2 &= a_0^3, \\ a_3 &= a_0^4, \\ a_4 &= a_0^5, \\ \cdots & \quad \cdots \end{aligned}$$

and that therefore (246) must have the form

$$y = a_0(1 + a_0 x + a_0^2 x^2 + \cdots).$$

This series is instantly recognized as having the sum

$$y = \frac{a_0}{1 - a_0 x},$$

which is identical with (245) provided $a_0 = -1/\alpha$.

As a second example, consider the differential equation

$$3y \frac{d^2y}{dx^2} + \left(\frac{dy}{dx}\right)^2 + 4y \frac{dy}{dx} + y^2 - 1 = 0, \tag{249}$$

which arises in certain fundamental studies connected with the vacuum tube. The circumstances of the problem are such that y must vanish and $\frac{dy}{dx}$ must equal unity when x is zero. Hence in this case a *particular* solution, not a general one, is desired.

Assume, then, that

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots$$

is a solution. If it is to vanish at $x = 0$, a_0 must be zero; and if $\frac{dy}{dx}$ is to reduce to unity, a_1 must be 1. Hence the series must become

$$y = x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \cdots.$$

Its derivatives are obviously

$$\frac{dy}{dx} = 1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \cdots,$$

$$\frac{d^2y}{dx^2} = 2a_2 + 6a_3 x + 12a_4 x^2 + \cdots,$$

so that (249) becomes

$$\begin{array}{r|rrrrl}
3y\dfrac{d^2y}{dx^2} & & +\,6a_2 x & +\,18a_3 x^2 & +\;36a_4 x^3 & \\
 & & & +\,6a_2^2 x^2 & +\,24a_2a_3 x^3 & +\cdots \\
+\left(\dfrac{dy}{dx}\right)^2 & +\,1 & +\,4a_2 x & +\,6a_3 x^2 & +\;8a_4 x^3 & \\
 & & & +\,4a_2^2 x^2 & +\,12a_2a_3 x^3 & +\cdots \\
+\,4y\dfrac{dy}{dx} & & 4x & +\,12a_2 x^2 & +\;16a_3 x^3 & \\
 & & & & +\;8a_2^2 x^3 & +\cdots \\
+\quad y^2 & & & +\quad x^2 & +\;2a_2 x^3 & +\cdots \\
-\,1 & -\,1 & & & & \qquad = 0,
\end{array}$$

and therefore leads to the equations

$$10a_2 + 4 = 0,$$

$$24a_3 + 10a_2^2 + 12a_2 + 1 = 0,$$

$$44a_4 + 36a_2a_3 + 16a_3 + 8a_2^2 + 2a_2 = 0.$$

From these it is found that

$$a_2 = -\tfrac{2}{5}, \quad a_3 = \tfrac{11}{120}, \quad a_4 = -\tfrac{47}{3300}, \quad \cdots.$$

The equation thus found is

$$y = x - \tfrac{2}{5}x^2 + \tfrac{11}{120}x^3 - \tfrac{47}{3300}x^4 + \cdots,$$

and as it converges quite rapidly it affords a pretty satisfactory solution of the problem with which it deals. No better solution is known.

There is another method of obtaining series solutions which is sometimes simpler than the one outlined above. We may illustrate it by the use of equation (244), which we have already solved subject to the boundary condition that $y = a_0$ at $x = 0$.

We know that the Taylor's series for any function $y(x)$ is

$$y(x) = y_0 + \frac{y_0'}{1!}x + \frac{y_0''}{2!}x^2 + \frac{y_0'''}{3!}x^3 + \cdots.$$

the y_0, y_0', $\cdots$, representing the value of y and its various derivatives at $x = 0$. We already know the first, y_0, from our boundary condition. The second, y_0', we may find at once by noting that, since (244) must be true at $x = 0$ as well as elsewhere, the derivative y_0' must be equal to y_0^2 and therefore to a_0^2. As for y_0'', we may find it by differentiating (244), thus getting

$$\frac{d^2y}{dx^2} = 2y\frac{dy}{dx},$$

and then inserting the values which we know y and $\dfrac{dy}{dx}$ take at $x = 0$. This gives

$$y_0'' = 2a_0^3.$$

Similarly,

$$\frac{d^3y}{dx^3} = 2\left(\frac{dy}{dx}\right)^2 + 2y\,\frac{d^2y}{dx^2}$$

gives

$$y_0''' = 6a_0^4,$$

and so on.

Substituting these in our general formula for the Taylor's series, we obtain

$$y(x) = a_0 + \frac{a_0^2\,x}{1!} + \frac{2a_0^3\,x^2}{2!} + \frac{6a_0^4\,x^3}{3!} + \cdots.$$

This, of course, reduces to

$$y = a_0(1 + a_0\,x + a_0^2\,x^2 + a_0^3\,x^3 + \cdots) = \frac{a_0}{1 - a_0\,x},$$

as before.

This scheme for finding the coefficients in the series often requires less labor than the method first explained; but it cannot always be applied. The reader will find, for example, that it does not lend itself readily to the solution of (249).

§ 78. *Pitfalls Connected with the Use of Series*

It must not be inferred, however, that a satisfactory power series can always be found. This is far from being the case.

For example, the equation

$$2xy\,\frac{dy}{dx} = 2x^3 + y^2$$

has as its solution

$$y = \sqrt{ax + x^3},$$

a function which cannot be expanded in a series of the type (246). If we attempt to get such a series we are led to the conclusion that all its coefficients must be zero.

Similarly, the equation

$$\frac{dy}{dx} = \frac{y}{x} + \frac{y}{x^2} - 1$$

has as its solution

$$y = \alpha x\,e^{-\frac{1}{x}} - x\,e^{-\frac{1}{x}}\int \frac{e^{\frac{1}{x}}}{x}\,dx;$$

but upon assuming a series solution and determining the coefficients the result

$$y = x^2 (1 + 1!\,x + 2!\,x^2 + 3!\,x^3 + \cdots)$$

is obtained This series *diverges for every value of* x, and is therefore useless for most purposes.

In other words, it is possible to get no answer, as in our first example, or to get a useless answer, as in our second. What is worse, it is *possible*, though it is unusual, to get a *wrong* answer. This, of course, is not a valid reason for shunning series solutions, or even for insisting upon assurance in advance that they will lead to correct results; but it does make extreme caution imperative when such assurance is not available.

§ 79. *Bessel's Equation*

Bessel's Equation,

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} + \left(1 - \frac{n^2}{r^2}\right) R = 0, \tag{250}$$

the solutions of which are known as Bessel Functions, arises in many problems of mathematical physics wherein it is convenient or necessary, because of boundary conditions, to use cylindrical coordinates. For example, the reader has met it in Problem 6, § 46.

Let us assume as a solution of (250) the power series

$$R = a_0 + a_1 r + a_2 r^2 + \cdots + a_j r^j + \cdots. \tag{251}$$

Then

$$\begin{aligned}
\frac{d^2R}{dr^2} &= && && 1\cdot 2a_2 + 2\cdot 3a_3 r + \cdots, \\
\frac{1}{r}\frac{dR}{dr} &= && a_1\frac{1}{r} + {} && 2a_2 + 3a_3 r + \cdots, \\
R &= && && a_0 + a_1 r + \cdots, \\
-\frac{n^2}{r^2} R &= -n^2 a_0 \frac{1}{r^2} && - n^2 a_1 \frac{1}{r} && - n^2 a_2 - n^2 a_3 r + \cdots.
\end{aligned}$$

If (251) is to be solution of (250) the combined sum of all the right-hand members of these equations must be equal to zero. Furthermore, if this sum is to equal zero for all values of r, the sum of the coefficients of like powers of r must be separately equal to zero. Therefore $n^2 a_0 = 0$, and if n is different from zero

$$a_0 = 0.$$

From the second column above we obtain $(1 - n^2)\, a_1 = 0$. Hence unless n is unity,

$$a_1 = 0.$$

From the next column we find $(4 - n^2)\, a_2 = 0$, whence, unless n is 2,

$$a_2 = 0.$$

Proceeding in this way, we can conclude that every a is zero up to a_n. The column in which a_n first appears, however, reduces to $(n^2 - n^2)\, a_n$ and is zero identically, no matter what the value of a_n may be. Hence the series (251) starts with the term in the nth power of r, the coefficient of which may have any value whatsoever.

Now consider the jth column, where $j > n$. It yields the equation

$$a_j (n^2 - j^2) = a_{j-2}.$$

From this it is immediately obvious that, since $a_{n-1} = 0$,

$$a_{n+1}, \quad a_{n+3}, \quad a_{n+5}, \quad \cdots,$$

are also zero. Also, it follows that

$$a_{n+2} = -\frac{a_n}{2(2n+2)},$$

$$a_{n+4} = \frac{a_n}{2\cdot 4(2n+2)(2n+4)},$$

$$a_{n+6} = -\frac{a_n}{2\cdot 4\cdot 6(2n+2)(2n+4)(2n+6)},$$

and so on.

We have now determined all the coefficients in the series (251) except a_n, which is a factor common to all the terms and is arbitrary. It is customary to replace it by a new arbitrary constant, A, such that

$$a_n = \frac{A}{2^n n!}.$$

Then it is easy to write all the coefficients in terms of this new A. In fact, we can deduce the general formula

$$a_{n+2p} = \frac{(-1)^p}{2^{n+2p}} \frac{A}{(n+p)!\,p!}.$$

A solution of (250) is therefore

$$R = A\left[\left(\frac{r}{2}\right)^n \frac{1}{n!} - \left(\frac{r}{2}\right)^{n+2} \frac{1}{(n+1)!\,1!} + \left(\frac{r}{2}\right)^{n+4} \frac{1}{(n+2)!\,2!} - \cdots\right]. \quad (252)$$

The bracketed term is ordinarily denoted by $J_n(r)$. It is known as the *Bessel function of the first kind and* nth *order*. It may also be written in the form

$$J_n(r) = \left(\frac{r}{2}\right)^n \sum_{p=0}^{\infty} \frac{\left(\frac{ir}{2}\right)^{2p}}{(n+p)!\,p!}. \quad (253)$$

The discussion given above was based on the assumption that n is an integer. The method for fractional values of n is similar. If n is fractional a series which satisfies (250) can be found in the form

$$R = a_n r^n + a_{n+1} r^{n+1} + \cdots,$$

in which, since n is fractional, all the powers are fractional. Physical problems sometimes lead to such fractional values of n, but they are not nearly so common in technical literature as are those involving integral powers. We need not discuss them further, except to remark that the bracketed term of (252), or (253), is still a valid definition of $J_n(r)$ even when n is

fractional, provided the factorials of the fractional quantities $n, n+1, \cdots$, are properly interpreted.

Finally, some problems of mathematical physics, notably that of determining the effective resistance of a conductor carrying a high-frequency current, lead to equations identical with (250), except that r is replaced by a new variable $\sqrt{-i}\, q$, in which q is a real quantity. Even in such cases, however, (253) is still a valid solution. If, then, we write $\sqrt{-i}\, q$ wherever r appears in (253), the right-hand side of this equation becomes complex. It is customary to think of the real and imaginary parts of this equation as separated from one another, and to denote them by the symbols $\mathrm{ber}_n\, q$ and $\mathrm{bei}_n\, q$, respectively.[1] In other words, $\mathrm{ber}_n\, q$ and $i\, \mathrm{bei}_n\, q$ are the real and imaginary parts of $J_n(\sqrt{-i}\, q)$:

$$J_n(\sqrt{-i}\, q) = \mathrm{ber}_n\, q + i\, \mathrm{bei}_n\, q.$$

Equation (250) is a second order differential equation and its complete solution should, therefore, involve two arbitrary constants. The solution given in (252) involves only one, and is therefore not the *general* solution of the problem. It can be shown that, when n is *not* an integer, the general solution is

$$R = A\, J_n(r) + B\, J_{-n}(r),$$

where $J_{-n}(r)$ is defined by merely writing $-n$ in place of n wherever it occurs in (252). On the other hand, when n *is* an integer the $J_{-n}(r)$ thus defined proves to be either equal to, or the negative of, $J_n(r)$, so that

$$R = [A \pm B]\, J_n(r) = C\, J_n(r).$$

As this contains only one arbitrary constant, there must be still another solution for (250) when n is an integer. This solution, for which a series can be found by a method into which we need not inquire, is ordinarily denoted by $K_n(r)$ and is called the *Bessel function of the second kind and* nth *order*. An important

[1] The *be* in these names is intended to suggest "Bessel," while the *r* and *i* terminations, respectively, denote "real" and "imaginary."

property of these functions of the second kind is that they all become infinite at $r = 0$. Since the boundary conditions imposed by most physical problems do not permit infinite values, the coefficient of B in the general solution

$$R = A\,J_n(r) + B\,K_n(r)$$

is usually found to be zero, which is tantamount to saying that the Bessel function of the second kind is not involved in the solution of the problem.

Extensive tables have been computed for Bessel functions of both the first and second kinds, as well as for the functions ber q and bei q to which we have referred. They are used in just the same manner as are the ordinary trigonometric or logarithmic tables.

§ 80. *Depressing the Order of a Linear Equation*

If, in attempting to solve a linear equation of order s, we know a particular solution *due to zero*, it is possible by its use to reduce the equation to a new one of order $s - 1$.

To see this, suppose $y = \phi(x)$ to be a particular solution of (134) *due to zero*. Let us replace y by the new variable z defined by the equation

$$y = \phi(x)\,z.$$

Then

$$\left.\begin{aligned} \frac{dy}{dx} &= z\frac{d\phi}{dx} + \phi\frac{dz}{dx}, \\ \frac{d^2y}{dx^2} &= z\frac{d^2\phi}{dx^2} + 2\frac{dz}{dx}\frac{d\phi}{dx} + \phi\frac{d^2z}{dx^2}, \\ \cdots \quad & \quad \cdots \quad \cdots \quad \cdots \end{aligned}\right\} \qquad (254)$$

and so on. All of these equations are linear in z and its derivatives. Hence, when we substitute them in (134) we shall get a new *linear* equation in z.

What is more, the definition of any derivative $\dfrac{d^jy}{dx^j}$ in (254) begins with the term $z\dfrac{d^j\phi}{dx^j}$, *and this is the only term in which* z

itself appears, rather than its derivatives. Hence, in our new differential equation for z, the coefficient of z itself must be $f_s(x)\dfrac{d^s\phi}{dx^s} + f_{s-1}(x)\dfrac{d^{s-1}\phi}{dx^{s-1}} + \cdots + f_0(x)\,\phi$; that is, $F_s\left(x, \dfrac{d}{dx}\right)\phi$. However, by definition, this is zero. In other words, *the new equation involves the derivatives of* z *up to the* sth, *but not* z *itself*. It is therefore of order $s - 1$ in the variable $z' = \dfrac{dz}{dx}$.

As an example of this process, let us take the equation

$$\frac{d^2y}{dx^2} + y = 0, \tag{255}$$

one solution of which is $y = \sin x$. If we write

$$y = z \sin x,$$

we find

$$\frac{d^2y}{dx^2} = -z \sin x + 2 \cos x \frac{dz}{dx} + \sin x \frac{d^2z}{dx^2},$$

whence

$$\sin x \frac{d^2z}{dx^2} + 2 \cos x \frac{dz}{dx} = 0,$$

or

$$\frac{dz'}{z'} + \frac{2 \cos x \, dx}{\sin x} = 0.$$

The solution of this is obviously

$$z' \sin^2 x = \alpha,$$

or

$$dz = \alpha \csc^2 x \, dx.$$

From this we find

$$z = \beta - \alpha \cot x,$$

or

$$y = \beta \sin x - \alpha \cos x.$$

This is indeed the *general* solution of (255).

As a second example, let us apply our process to Bessel's equation (250), *one* solution of which is known to be $J_n(r)$. We write

$$R = zJ_n(r),$$

$$\frac{dR}{dr} = z\frac{dJ_n}{dr} + J_n\frac{dz}{dr},$$

$$\frac{d^2R}{dr^2} = z\frac{d^2J_n}{dr^2} + 2\frac{dz}{dr}\frac{dJ_n}{dr} + J_n\frac{d^2z}{dr^2};$$

whence (250) becomes

$$J_n\frac{d^2z}{dr^2} + \left(2\frac{dJ_n}{dr} + \frac{1}{r}J_n\right)\frac{dz}{dr} + \left[\frac{d^2J_n}{dr^2} + \frac{1}{r}\frac{dJ_n}{dr} + \left(1 - \frac{n^2}{r^2}\right)J_n\right]z = 0.$$

However, we know that the coefficient of z is zero, since J_n is a solution of (250). Hence this equation reduces to

$$J_n\frac{dz'}{dr} + \left(2\frac{dJ_n}{dr} + \frac{1}{r}J_n\right)z' = 0,$$

or

$$\frac{dz'}{z'} + 2\frac{dJ_n}{J_n} + \frac{dr}{r} = 0,$$

the solution of which is

$$z'rJ_n^2 = \alpha,$$

or

$$z = \beta + \alpha\int\frac{dr}{rJ_n^2(r)}.$$

Except for difficulties of integration, this is an entirely acceptable *general* solution of (250). From it, indeed, we could get the function $K_n(x)$ to which we referred in § 79; but as it happens not to be the simplest way to derive it, it is perhaps just as well to refrain.

The important point to note is, that in case a *particular* solution of a second order equation is known, the equation can always be replaced by another of the first order, and thus its *general* solution found.

ANSWERS TO THE PROBLEMS

ANSWERS TO THE PROBLEMS

CHAPTER I, § 5, PAGE 19.

1. $\frac{1}{x}\frac{d^2w}{dx^2}+\frac{w}{x}.$

2. $x^2\frac{d^2w}{dx^2}+4x\frac{dw}{dx}+2w=0; \quad y=\left(\frac{\alpha}{x}+\frac{\beta}{x^2}\right)^2.$

3. $\frac{dw}{dx}=a\,e^{cx}; \quad y=\frac{a}{c}+\alpha\,e^{-cx}.$

4. $\frac{d^2\theta}{dt^2}+k^2\theta\,e^{2t}=0.$

5. $(1-x^2)\frac{d^2P}{dx^2}-2x\frac{dP}{dx}+m\,P=0.$

6. $\frac{dw}{w}=\frac{dx}{x}.$

7. $\frac{dy}{dx}=2xy.$

8. $\frac{dz}{dx}=e^{-(1+n)w}\left(\frac{d^{m+1}y}{dw^{m+1}}-n\frac{d^my}{dw^m}\right).$

11. $(x-\cos\theta)^2+(y-\sin\theta)^2=r^2.$

12. $x^2+y^2=(1\pm r)^2.$

13. $y=-x^2/8.$

14. $x=\sin\theta-\cos^3\theta; \quad y=\sin^3\theta-\cos\theta.$

15. $3\pi/32.$

16. (*a*) Ordinary, 2nd order, 3rd degree, non-linear.
 (*b*) Ordinary, 1st order, 1st degree, linear with variable coefficients.
 (*c*) Partial, 2nd order, 1st degree, linear with constant coefficients.

17. (*a*) Ordinary, 2nd order, 1st degree, non-linear.
 (*b*) Ordinary, 2nd order, 1st degree, non-linear unless $f(y, z)$ is linear.

18. $\frac{dw}{du}+2\left(\frac{1}{u}-1\right)w=3\,e^{2u}.$

CHAPTER II, § 11, PAGE 37.

1. $\alpha = \dfrac{\theta'_0 \cos pt_1 - \theta'_1 \cos pt_0}{p \sin p(t_1 - t_0)}; \quad \beta = \dfrac{\theta'_0 \sin pt_1 - \theta'_1 \sin pt_0}{p \sin p(t_1 - t_0)}.$

2. $\alpha = \dfrac{\theta_0 p \cos pt_1 - \theta'_1 \sin pt_0}{p \cos p(t_0 - t_1)}; \quad \beta = \dfrac{\theta_0 p \sin pt_1 + \theta'_1 \cos pt_0}{p \cos p(t_0 - t_1)}.$

CHAPTER III, § 20, PAGE 56.

4. $\dfrac{d\theta}{dt} = \kappa(\theta_0 - \theta)$ where θ is the reading of the thermometer and θ_0 is the temperature of the cold medium.

5. $\dfrac{d^2y}{dx^2} \cdot \dfrac{dy}{dx} = k\left[1 + \left(\dfrac{dy}{dx}\right)^2\right]^{\frac{3}{2}}.$

6. $y^2\left(\dfrac{dy}{dx}\right)^2 + y^2 = r^2.$

7. $\dfrac{8}{27}\left(\dfrac{dy}{dx}\right)^3 + \dfrac{4}{9}\left(\dfrac{dy}{dx}\right)^2 = x + y.$

8. $4y\left(\dfrac{dy}{dx}\right)^3 - 2x^2\left(\dfrac{dy}{dx}\right)^2 + 4xy\left(\dfrac{dy}{dx}\right) + x^3 = 16y^2.$

9. $\dfrac{dy}{dx} + \dfrac{y}{\sqrt{L^2 - y^2}} = 0.$

10. $\dfrac{dy}{dx} = ky\,(x - 2).$

11. $m\dfrac{dv}{dt} = F - \dfrac{v}{t + 1}.$

CHAPTER IV, § 23, PAGE 62.

1. $\tan\theta = \alpha + ms/\kappa; \quad \alpha = 0.$

2. $y = \alpha e^{-cx} + a/c.$

3. $\theta = \theta_0 + \alpha e^{-\kappa t}$; 0.520° higher than the surrounding medium; fair.

4. (*a*) $\sec x = \alpha - \tan y.$

 (*b*) $y\sqrt{1 - x^2} + x\sqrt{1 - y^2} = \alpha.$

 (*c*) $y = b + \alpha x/(bx + 1).$

5. $x = \dfrac{M e^{(Nm - Mn)(\alpha + kt)} - N}{m e^{(Nm - Mn)(\alpha + kt)} - n}.$

6. $y + \tan^{-1} y = \alpha + x + \log x.$

CHAPTER IV, § 25, PAGE 69.

1. $y = e^x\left(\alpha + \log x - x + \dfrac{x^2}{2.2!} - \dfrac{x^3}{3.3!} + \dfrac{x^4}{4.4!} - \cdots\right).$

2. $y = \alpha e^x - \dfrac{1}{x} + \dfrac{1}{x^2} - \dfrac{2!}{x^3} + \dfrac{3!}{x^4} - \dfrac{4!}{x^5} + \cdots .$

3. $\mathrm{Ci}\ 1 = 0.3374; \quad \mathrm{Ci}\ 10 = -0.0455.$

CHAPTER IV, § 29, PAGE 81.

1. $y = \alpha e^{y/x}.$

2. $y = \alpha x\sqrt{x^2 + y^2}.$

3. $ax^2 + byx + cy^2 - gx + ey = \alpha.$

4. $\sin^2 x + \sin^2 y = \alpha.$

5. $2t = mI + I\sqrt{4 - m^2}\tan\left[\sqrt{4 - m^2}\left(\log \alpha\sqrt{I^2 - mtI + t^2}\right)/m\right].$

6. $r^2 - \theta^2 = \alpha r^3.$

7. $\cos(a\alpha + b\beta) + \sin(b\alpha + a\beta) = \gamma.$

8. $T^2 + t^2 + 2\sin^{-1}(T/t) = \alpha.$

CHAPTER IV, § 31, PAGE 85.

1. $y = 1 + \alpha e^{-x^2/2}.$

2. $xy = \alpha - x\cos x + \sin x.$

3. $x^4y^4 = \alpha - 4x^4\cos x + 16x^3\sin x + 48x^2\cos x - 96x\sin x - 96\cos x.$

4. $\rho = a\theta + \alpha\theta\sqrt{1 - \theta^2}.$

5. $T = (1 + \alpha t + \log t)^{-1}.$

6. $y = \alpha e^{-\sin x} + \sin x - 1.$

7. $y = \dfrac{\sec x + \tan x}{\sin x + \alpha},$

8. $y = \alpha e^{-\phi} + \phi - 1.$

9. $\theta = \theta_0 + \alpha e^{-\kappa t}.$

CHAPTER IV, § 32, PAGE 89.

2. $x + y + 2 \mp 2\sqrt{1 + xy} = \alpha + 2\log\left(-1 \pm \sqrt{1 + xy}\right) - 2\log x.$

3. $\phi = \frac{1}{27} - \frac{1}{9}(\alpha + 3r) \pm \frac{2}{27}(\alpha + 3r)^{3/2}.$

4. $z + 1 = \log\left(-1 \pm \sqrt{\alpha + 2y}\right) \pm \sqrt{\alpha + 2y}.$

5. $\alpha\left(2T - t\sqrt{t^2 + T}\right)^{\sqrt{17}} = \dfrac{4\sqrt{t^2 + T} - t(1 - \sqrt{17})}{4\sqrt{t^2 + T} - t(1 + \sqrt{17})}.$

6. $y = \alpha x + \phi(\alpha).$

7. $e^y = \alpha\left(x + \sqrt{x^2 - 1}\right).$

CHAPTER IV, § 33, PAGE 91.

1. $\theta = \alpha \sin(pt + \beta)$.

2. $\tan\theta + \alpha = ms/\kappa$.

3. $2m(\beta + y) = \kappa\,(e^{m(x+\alpha)/\kappa} + e^{-m(x+\alpha)/\kappa})$.

4. $ne = \sqrt{2e/m}\,\phi^{3/2}/9\pi x^2$.

5. $12\pi nex + \alpha$
$= \sqrt{\beta + 8\pi mn\sqrt{v_0^2 + 2e\,(\phi - V_0)/m}}\left(\sqrt{v_0^2 + 2e\,(\phi - V_0)/m} - \beta/4\pi mn\right)$.

6. $cpf(x) = \alpha \sin(cpx + \beta)$.

7. $\beta + ky = \sqrt{1 - (kx + \alpha)^2} + \log(kx + \alpha) - \log\left(1 + \sqrt{1 - (kx + \alpha)^2}\right)$.

9. (a) $v = \alpha e^{3u}/u^2$.
(b) $\sin^{-1} v = \alpha - 2\sqrt{1 - u^2}$.
(c) $v = \alpha - u + e^{2u}/8$.
(d) $\log y = \alpha - \frac{1}{2}\cos(x^2 - 1) - \frac{4}{3}x^{3/2}$.
(e) $\alpha\sqrt{x^2 + y^2} = e^{\tan^{-1} y/x}$.

10. (a) $v = 1 + \alpha e^{-u^2}$.
(b) $\log(1 + v^2) + 2\tan^{-1} u = \alpha$.
(c) $\tan v = \alpha + \log\log u$.

11. $x = \alpha + L\log y - L\log(L - \sqrt{L^2 - y^2}) - \sqrt{L^2 - y^2}$.

12. $2\log y = \alpha + k\,(x - 2)^2$.

13. $(m + 1)\,v = F[t + 1 - 1/(t + 1)^{1/m}]$.

CHAPTER V, § 36, PAGE 99.

1. $(27y^2 - 2x^3)(16^2y^4 - 32x^3y^2 + x^6) = 0$.

2. $\rho = \frac{1}{2}$.

3. $\left(\dfrac{d\rho}{d\theta}\right)^2 = \rho^2(2\rho - 1);\quad \rho = \frac{1}{2}$.

CHAPTER VI, § 39, PAGE 105.

1. $\theta = 4°\text{ C.} + 0.126\,S/a$.

2. $\theta = 52.9°\text{ C.}$

3. The thicker wire.

4. $pH = 2\pi rg\,(\theta_0 + \theta_1)\tanh(pl/2)$.

5. $\theta_l = kp\theta_0/(kp\cosh pl + g\sinh pl)$.

CHAPTER VI, § 41, PAGE 110.

1. $\log \beta(x^2 + y^2) = -2 \tan \alpha \tan^{-1}(y/x)$.

2. $x = \beta y$.

3. $\sqrt{8 \tan^2 \alpha - 1} \log \beta(y^2 \tan \alpha + 2x^2 \tan \alpha - xy)$
$$= 6 \tan^{-1}\left[(2y \tan \alpha - x)/x\sqrt{8 \tan^2 \alpha - 1}\right].$$

4. $(1 - \beta^2)^{3/2}(y + \gamma) = \beta\sqrt{1 - \beta^2}\sqrt{x^2(1 - \beta^2) + 2\alpha\beta x - \alpha^2}$
$$- \alpha \log\left(\sqrt{x^2(1 - \beta^2) + 2\alpha\beta x - \alpha^2} + x\sqrt{1 - \beta^2} + \alpha\beta/\sqrt{1 - \beta^2}\right).$$

5. $y^2 + z^2 = \alpha + 2 \log \cos x$.

CHAPTER VI, § 44, PAGE 118.

1. $24\, qy = w\,(x^4 - 4\, lx^3 + 6\, l^2x^2)$, if w is the load per unit length and the beam is held horizontal at $x = 0$.

2. $12\, qy = W(3\, lx^2 - 2|x|^3)$, where x is measured from the knife edge.

3. $x = \beta + \displaystyle\int \frac{[W(y + \epsilon)^2 - \alpha]\, dy}{\sqrt{4q^2 - [W(y + \epsilon)^2 - \alpha]^2}}$.

4. $\dfrac{y + \epsilon}{y_0 + \epsilon} = \sin\left[\dfrac{2x}{l}\left(\sin^{-1}\dfrac{\epsilon}{y_0 + \epsilon}\right) + \dfrac{\pi}{2}\left(1 - \dfrac{2x}{l}\right)\right]$, where ϵ is the length of the top and bottom arms, and y_0 is the deflection of the mid-point of the vertical beam.

CHAPTER VI, § 46, PAGE 124.

1. $f(x) = A \sin n\pi x$, where n is an integer.

3. $k(t) = B \sin\,(n\pi t\sqrt{T/m} - C)$; $\dfrac{n}{2}\sqrt{T/m}$.

4. They are proportional to the integers 1, 2, 3, $\cdots$.

5. The amplitude and phase of the vibration.

6. $c^2 \dfrac{\partial^2 x}{\partial t^2} = \dfrac{\partial^2 x}{\partial r^2} + \dfrac{1}{r}\dfrac{\partial x}{\partial r}$.

CHAPTER VI, § 51, PAGE 144.

1. $(y_1 - y_0)/(x_1 - x_0) = (y - y_0)/(x - x_0)$.

2. $\sinh^{-1}(1.007\, y) = \cosh^{-1}(1.007\, x) - 0.125$.

3. $x = t$.

4. A cylinder.

5. $r = \sqrt[3]{15\, V/4\pi}\, \sqrt{\cos \theta}$.

6. $y = 1 + \alpha \cosh(x/\alpha) - \alpha \cosh(1/\alpha)$, where α is a root of the equation $2\alpha \sinh(1/\alpha) = 3$. This equation has two roots, $\alpha = \pm 0.6165$. The negative one gives a surface of *maximum* area.

7. A sphere.

8. $\phi = \pi/2$.

9. $x = \beta + \int \frac{\lambda\, dy}{\sqrt{[\alpha - \log(1+y)]^2 - \lambda^2}}$.

Chapter VII, § 56, Page 163.

1. $\theta = \alpha_1 e^{px} + \alpha_2 e^{-px}$.
2. $y = \alpha_1 e^{ix} + \alpha_2 e^{-ix}$.
3. $y = \alpha_1 e^{3x} + \alpha_2 e^{4x}$.
4. $r = \alpha_1 e^{a\theta} + \alpha_2 e^{-a\theta}$.
5. $y = \alpha_1 e^{ax} + \alpha_2 e^{-ax} + \alpha_3 e^{iax} + \alpha_4 e^{-iax}$.
6. $v = \frac{1}{29} e^{-2u} + \alpha_1 e^{(3+2i)u} + \alpha_2 e^{(3-2i)u}$.
7. $y = \alpha_1 e^{(-2+\sqrt{5})t} + \alpha_2 e^{(-2-\sqrt{5})t} - \frac{1}{10}\sin t - \frac{1}{5}\cos t$.
8. $y = \alpha_1 e^{\sqrt{3}\,ix} + \alpha_2 e^{-\sqrt{3}\,ix} + \frac{1}{2}\sin x - \frac{1}{18}\sin 3x$.
9. $x = \alpha_1 e^{\sqrt{2}\,(1-i)t} + \alpha_2 e^{-\sqrt{2}\,(1-i)t} + \alpha_3 e^{\sqrt{2}\,(1+i)t} + \alpha_4 e^{-\sqrt{2}\,(1+i)t} + \frac{1}{17} e^{it}$.
10. $x = \alpha_1 e^{-t/5} + \frac{1}{226}\sin 3t - \frac{15}{226}\cos 3t$.
11. $x = \alpha_0 + \alpha_1 e^{t} + \alpha_2 e^{2t} + \alpha_3 e^{3t} + \frac{1}{360} e^{-3t}$.
14. $y = \alpha_1 x^6 + \alpha_2 x^2 + \alpha_3 + \alpha_4 x^{-3} + \frac{17}{216} x^6 \log x$.
15. $x = (\alpha_1 + \alpha_2 \log t)/t + \alpha_3 t^2 + \alpha_4 t^3 + \frac{1}{47450}[62\cos(3\log t) + 141\sin(3\log t)]$.

Chapter VII, § 58, Page 175.

1. $I = \frac{\alpha}{RC} e^{-t/CR} + E_0 \frac{R\cos nt - \frac{1}{Cn}\sin nt}{\frac{1}{C^2n^2} + R^2}$.

2. $I = \frac{\alpha R}{L} e^{-Rt/L} + E_0 \frac{R\sin nt - Ln\cos nt}{R^2 + L^2n^2}$; Zero.

3. $I = \alpha_1 \cos\frac{t}{\sqrt{LC}} + \alpha_2 \sin\frac{t}{\sqrt{LC}} + \frac{E_0 Cn\cos nt}{1 - LCn^2}$; $\frac{1}{2\pi}\sqrt{\frac{1}{LC}}$; An alternating current of frequency $\frac{1}{2\pi}\sqrt{\frac{1}{LC}}$; Indefinitely.

4. $I = -\frac{E_0 Cn}{1 - LCn^2}\cos\frac{t}{\sqrt{LC}} + \frac{E_0 Cn}{1 - LCn^2}\cos nt$. The steady state is never reached.

5. $1/2\pi\sqrt{LC}$; No.

6. The natural frequency.

7. $\bar{I} = \frac{E_0}{Ln} e^{-Rt/2L} \sin \overline{nt}.$

8. $R^2 I = 0,$ $\quad t < 0;$
$R^2 I = Rt - L(1 - e^{-Rt/L}),$ $\quad 0 < t < 1;$
$R^2 I = -R(t-2) + L(1 - 2e^{-R(t-1)/L} + e^{-Rt/L}),$ $\quad 1 < t < 2;$
$R^2 I = L(e^{-R(t-2)/L} - 2e^{-R(t-1)/L} + e^{-Rt/L}),$ $\quad 2 < t.$

Chapter VII, § 61, Page 189.

1. $y = \alpha_1 e^{-x} + \alpha_2 e^{-2x} + \frac{1}{2}x^2 - \frac{3}{2}x + \frac{7}{4};$ Yes.

2. (a) $a^2 y = \alpha e^{ax} - ax - 1.$
(b) $a^{n+1} y = \alpha e^{ax} - [(ax)^n + n(ax)^{n-1} + n(n-1)(ax)^{n-2} + \cdots + n!].$
(c) $(a^2 + n^2) y = \alpha e^{ax} - a \sin nx - n \cos nx.$
(d) $(n - a) y = \alpha e^{ax} + e^{nx}.$

Chapter VII, § 63, Page 196.

1. $y = (\alpha_1 + \alpha_2 x) e^x + \alpha_3 e^{-x}.$

2. $y = \alpha_1 + (\alpha_2 + \alpha_3 x + \alpha_4 x^2) e^x + \frac{1}{2} e^{2x}.$

3. $r = \alpha_1 e^\theta + \alpha_2 \cos\theta + \alpha_3 \sin\theta - \frac{1}{4}(\cos\theta + \theta\cos\theta + \theta\sin\theta).$

8. $xy = \alpha_1 + \alpha_2 \log x + \frac{1}{2}(\log x)^2.$

Chapter VIII, § 74, Page 227.

1. (a) $y = 1 - 2\alpha_1 e^x - (4 - \sqrt{2}\,i)\alpha_2 e^{-(1+\sqrt{2}\,i)x/3} - (4 + \sqrt{2}\,i)\alpha_3 e^{-(1-\sqrt{2}\,i)x/3},$
$z = \alpha_1 e^x + \alpha_2 e^{-(1+\sqrt{2}\,i)x/3} + \alpha_3 e^{-(1-\sqrt{2}\,i)x/3}.$

(b) $I_1 = (\frac{1}{2}\alpha_2 + \alpha_3 + \alpha_3 t) e^{-t} + (\cos nt - n^2 \cos nt + 2n \sin nt)/(1 + n^2)^2,$
$I_2 = (\alpha_1 + \alpha_2 t + \alpha_3 t^2) e^{-t}$
$\qquad + 2n(\sin nt - 3n^2 \sin nt - 3n \cos nt + n^3 \cos nt)/(1 + n^2)^3.$

(c) $y = \alpha_1 e^x + 6\alpha_2 e^{(-3-\sqrt{3}\,i)x/2} + 6\alpha_3 e^{(-3+\sqrt{3}\,i)x/2},$
$z = \alpha_1 e^x + \alpha_2 (1 - \sqrt{3}\,i) e^{(-3-\sqrt{3}\,i)x/2} + \alpha_3 (1 + \sqrt{3}\,i) e^{(-3+\sqrt{3}\,i)x/2} - x/3.$

(d) $2y_1 = (2\alpha_1 + \alpha_2 + 2\alpha_2 x) e^{-4x} + e^{-4x}\int e^{4x}[f_1(x) + f_2(x)]\,dx$
$\qquad + e^{-4x}\int dx \int e^{4x}[6f_1(x) - 2f_2(x)]\,dx,$
$2y_2 = (\alpha_1 + \alpha_2 x) e^{-4x} + e^{-4x}\int e^{4x}[-f_1(x) + f_2(x)]\,dx$
$\qquad + e^{-4x}\int dx \int e^{4x}[3f_1(x) - f_2(x)]\,dx.$

$$2.\ 6y_1 = \alpha_1 e^{-x/3} + \alpha_2 e^{-3x} - \frac{df_1(x)}{dx} + \frac{10}{3} f_1(x) + f_2(x)$$
$$+ \frac{1}{36} e^{-x/3} \int e^{x/3} [23 f_1(x) + 15 f_2(x)]\, dx$$
$$- \tfrac{3}{4} e^{-3x} \int e^{3x} [21 f_1(x) + 5 f_2(x)]\, dx,$$

$$6y_2 = \frac{7}{2} \alpha_1 e^{-x/3} - \frac{1}{2} \alpha_2 e^{-3x} + \frac{df_1(x)}{dx} - \frac{10}{3} f_1(x) - f_2(x)$$
$$+ \tfrac{1}{72} e^{-x/3} \int e^{x/3} [161 f_1(x) + 105 f_2(x)]\, dx$$
$$+ \tfrac{1}{8} e^{-3x} \int e^{3x} [63 f_1(x) + 15 f_2(x)]\, dx.$$

$$3.\ 2y_1 = 2f_1(x) - 3f_2(x) + \frac{df_1(x)}{dx} - \frac{df_2(x)}{dx},$$

$$2y_2 = f_2(x) - \frac{df_1(x)}{dx} + \frac{df_2(x)}{dx}.$$

INDEX